सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI

24486

I.A.R I.6.
GIP NLK—H 3 I.A.R.I —10-5 55—15,000

PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 194

LONDON
Published for the Royal Society by the
Cambridge University Press
Bentley House, N.W.1

9 November 1948

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)
and published by the Cambridge University Press
(Cambridge, and Bentley House, London)
Agents for Canada and India: Macmillan

CONTENTS

SERIES A VOLUME 194

No. A 1036. 28 July 1948

PAGE

On the calculation of surface tension from measurements of pendant drops. By S. Fordham 1

Concerning the effect of compressibility on laminar boundary layers and their separation. By L. Howarth 16

Intensity measurements in the very soft X-ray region. By F. C. Chalklin . . 42

The intermediate state of superconductors. I. Magnetization of superconducting cylinders in transverse magnetic fields. By M. Désirant and D. Schoenberg . 63

The intermediate state of superconductors. II. The resistance of cylindrical superconductors in transverse magnetic fields. By E. R. Andrew 80

The intermediate state of superconductors. III. Theory of behaviour of superconducting cylinders in transverse magnetic fields. By E. R. Andrew . . 98

Aspect of a theory of unimolecular reaction rates. By N. B. Slater 112

The transmutation of magnesium, aluminium and silicon by deutrons. By H. R. Allan and Christiane A. Wilkinson 131

No. A 1037. 12 August 1948

Heat transfer by radiation to surfaces at low temperatures. By M. Blackman, Sir Alfred Egerton, Sec.R.S. and E. V. Truter 147

Spectroscopic studies of low-pressure flames; temperature measurements in acetylene flames. By A. G. Gaydon and H. G. Wolfhard 169

'Diffuse scattering' of the Fermi electrons in monovalent metals in relation to their electrical resistivities. By A. B. Bhatia and Sir K. S. Krishnan, F.R.S. 185

A generalization of the method of separating longitudinal and transverse waves in electrodynamics. By Sonja Ashauer 206

The laminar boundary-layer equations of bodies of revolution. Motion of a sphere. By D. Meksyn 218

On the scattering of protons by protons. By W. H. Ramsey 228

Determination of the viscosity of molten salts by a capillary method. The system cadmium chloride-cadmium bromide. By H. Bloom, B. S. Harrap and E. Haymann 237

A general kinetic theory of liquids. V. Liquid He II. By H. S. Green 244

Electroviscosity. I. The flow of liquids between surfaces in close proximity. By G. A. H. Elton 259

Electroviscosity. II. Experimental demonstration of the electroviscous effect. By G. A. H. Elton 275

No. A 1038. 2 September 1948

PAGE

The use of flat-ended projectiles for determining dynamic yield stress. I. Theoretical considerations. By Sir Geoffrey Taylor, F.R.S. 289

The use of flat-ended projectiles for determining dynamic yield stress. II. Tests on various metallic materials. By A. C. Whiffin 300

The use of flat-ended projectiles for determining dynamic yield stress. III. Changes in microstructure caused by deformation under impact at high-striking velocities. By W. E. Carrington and Marie L. V. Gayler. (Plates 1 to 8) . 323

The determination of the masses necessary to produce a given gravitational field. By E. C. Bullard, F.R.S. and R. I. B. Cooper 332

The velocity of propagation of electromagnetic waves derived from the resonant frequencies of a cylindrical cavity resonator. By L. Essen and A. C. Gordon-Smith 348

The constitution of the aluminium-rich aluminium-cobalt-iron alloys, with reference to the role of transitional elements in alloy formation. By G. V. Raynor and M. B. Waldron. (Plate 9) 362

A comparative X-ray study of foetal and adult sheep haemoglobins. By J. C. Kendrew and M. F. Perutz. (Plate 10) 375

Laboratory experiments on whirlpools. By A. M. Binnie and G. A. Hookings . . 398

No. A 1039. 9 November 1948

The kinetic salt effect in monolayer reactions. By J. T. Davies and E. K. Rideal, F.R.S. 417

Atmospheric tides in the ionosphere. III. Lunar tidal variations at Canberra. By D. F. Martyn 429

Atmospheric tides in the ionosphere. IV. Studies of the solar tide, and the location of the regions producing the diurnal magnetic variations. By D. F. Martyn 445

Temperature and pressure phenomena in the flow of saturated liquids. By R. S. Silver 464

The detonation of solid explosives: the equilibrium conditions in the detonation wave-front and the adiabatic expansion of the products of detonation. By H. Jones and A. R. Miller 480

A study of expected loss rates in the counting of particles from pulsed sources. By C. H. Westcott 508

Decay of turbulence in the final period. By G. K. Batchelor and A. A. Townsend 527

On the classical theory of particles. By C. Jayaratnam Eliezer 543

Adjustments in bacterial reaction systems. I. The reducing power of *Bact. lactis aerogenes* under various conditions. By P. R. Lewis and C. N. Hinshelwood, F.R.S. (Abstract) 556

Adjustments in bacterial reaction systems. II. Adaptive mechanisms. By C. N. Hinshelwood, F.R.S. and P. R. Lewis. (Abstract) 556

Index 557

On the calculation of surface tension from measurements of pendant drops

By S. Fordham

Research Department, I.C.I. Ltd., Explosives Division, Stevenston, Ayrshire

(*Communicated by F. A. Freeth, F.R.S.—Received* 13 *November* 1947—
Revised 15 *March* 1948)

The calculations presented are an extension of the work of Bashforth & Adams (1883) to give the shapes of pendant drops of liquids for values of β from $-0{\cdot}25$ to $-0{\cdot}6$ at intervals of 0·025. The results have been used to calculate, to an accuracy of 0·001 to 0·01 %, the constants needed for measuring surface and interfacial tensions by the method suggested by Andreas, Hauser & Tucker (1938), which has thus been made independent of calibration.

INTRODUCTION

Many methods have been proposed for the measurement of surface and interfacial tensions, and each may be said to have advantages for particular applications. Some are dynamic, and are ideally suited to following the changes occurring in a surface during the first fractions of a second after its formation. More commonly, however, measurements are made in a static manner, and are intended to indicate the properties of the surface in its equilibrium condition. The static methods all purport to measure the same property, but they may differ greatly in suitability for individual purposes. A valuable general discussion of these methods has been given by Adam (1938).

At the present time the basic method for determining surface tension involves the use of a capillary tube, by measuring either the capillary rise or else the hydrostatic pressure needed to restore the meniscus to its normal level. The results are calculated, following Sugden (1921), with the help of the tables of Bashforth & Adams (1883) for the shape of free surfaces of liquids. To obtain accuracy, however, extreme precautions are needed, and the method is not well adapted to routine use. A number of secondary methods have therefore been evolved, which are more rapid in operation. The maximum bubble pressure method of Sugden (1922) has a theoretical foundation, being based also on the tables of Bashforth & Adams. The drop-weight method of Harkins & Brown (1919) and the ring method of de Noüy (1919) are more empirical. All these techniques for measuring surface tension depend finally upon comparison with the capillary method using such liquids as benzene. Generally they involve enlarging and breaking the surface of the liquid, and therefore are not strictly static in nature.

Some time ago the author wished to carry out routine measurements of the surface tensions of some colloidal solutions. These solutions were characterized by relatively high viscosity, and by containing material of high molecular weight and thus of low speed of diffusion. A method was therefore required which did not depend on an extension of the surface, because under such conditions not only would the slow diffusion prevent the attainment of equilibrium, but also the high viscosity would make invalid any calibration with free-flowing liquids. A modification of the technique proposed by Wilhelmy (1863) was considered, in which the surface remained unchanged, but this was discarded in favour of a method more recently developed by Andreas, Hauser & Tucker (1938), based on measurements of the shape and size of pendant drops.

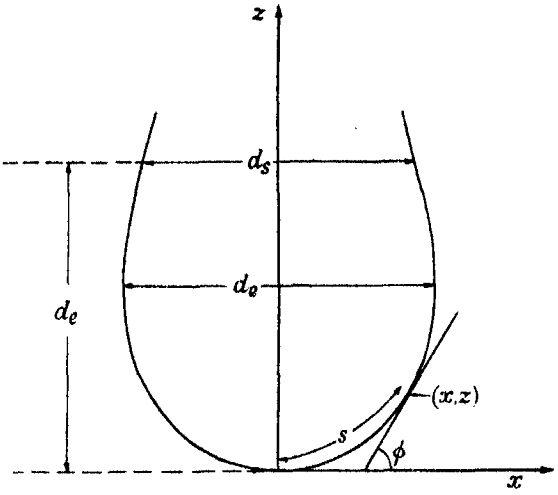


FIGURE 1. Pendant drop. $\beta = -0{\cdot}45$.

The use of pendant drops has been suggested by several workers, such as Worthington (1881, 1885) and Ferguson (1912), but proposals have usually involved difficult measurements, as of the position of points of inflexion on the profile of the drop. Andreas *et al.* suggested a simpler method of calculation. Taking the origin at the bottom of the drop, and co-ordinates x and z in a horizontal and vertical direction respectively, the equation for the profile of a liquid surface of revolution may be written as

$$\frac{1}{\rho}+\frac{\sin\phi}{x}=2+\beta z, \tag{1}$$

where ρ is the radius of curvature at the point (x, z), ϕ is the slope at the point (x, z), and

$$\beta = -g\sigma b^2/\gamma, \tag{2}$$

where σ is the density of the liquid, γ is the surface tension, and b is the radius of curvature at the origin. The unit of length adopted in equation (1) is b, so that x, z and β are non-dimensional. It will thus be seen that β determines the shape of a drop, and b its size. The negative sign for β for a pendant drop follows the convention adopted by Bashforth & Adams (1883).

Andreas *et al.* avoided the actual determination of β and b by a method, illustrated in figure 1, in which the maximum diameter d_e was measured, and also the diameter d_s at a horizontal plane distant d_e from the bottom of the drop. Denoting the ratio d_s/d_e by S, it is clear that S is a function of β, whilst for a given S, d_e is proportional to b. Thus, by (2),

$$\gamma = g\sigma d_e^2/H,$$

where

$$1/H = 1/\beta(d_e/b)^2 = 1/4\beta x_e^2 \tag{3}$$

is a function of $S = d_s/d_e$. This function, connecting $1/H$ and S, was obtained by experiments with water, the surface tension of which was known from capillary measurements.

There is no reason why the relation between S and $1/H$ should not be calculated, and the method thus placed on a firm theoretical foundation and made independent of experimental calibration. Although Andreas *et al.* realized this, they tended to doubt the accuracy of available mathematical methods; but with sufficient care, numerical integrations may be carried out to any finite degree of accuracy which may be required.

METHOD OF CALCULATION

At any point (x, z),

$$\frac{d\phi}{ds} = \frac{1}{\rho} \tag{4}$$

or, by equation (1),

$$\frac{d\phi}{ds} = 2 + \beta z - \frac{\sin\phi}{x}, \tag{5}$$

$$\frac{dx}{ds} = \cos\phi, \tag{6}$$

$$\frac{dz}{ds} = \sin\phi, \tag{7}$$

where s = the arc distance of the point (x, z) from the origin at the bottom of the drop. The unit of length in these equations is again b, the radius of curvature at the bottom of the drop.

For each value of β, simultaneous finite difference integration gives ϕ, x and z in terms of s, after the first values have been obtained by series expansions. At each step an estimate is made of the next entry in the finite difference table of $d\phi/ds$ against s, and the value of ϕ calculated by the integration formula (expressed in terms of backward differences from the end of the interval of integration)

$$\phi_s = \phi_{s-\delta s} + \delta s(\phi'_s - \tfrac{1}{2}\Delta\phi'_s - \tfrac{1}{12}\Delta^2\phi'_s - \tfrac{1}{24}\Delta^3\phi'_s - \tfrac{19}{720}\Delta^4\phi'_s - \tfrac{3}{160}\Delta^5\phi'_s). \tag{8}$$

This value of ϕ by equations (6) and (7) gives the next entries in the tables of dx/ds and dz/ds, so that a formula similar to (8) may be used to calculate x and z. These figures, substituted in equation (5), should give a value for $d\phi/ds$ equal to that assumed. The manner of correcting for differences between the calculated and assumed values of $d\phi/ds$ is given by Bashforth & Adams (1883), who also give tables of results for many values of β, of which $-0{\cdot}3$, $-0{\cdot}4$, $-0{\cdot}5$ and $-0{\cdot}6$ are relevant to the present problem.

This method was used to obtain tables for all other values of β between $-0{\cdot}25$ and $-0{\cdot}6$, at intervals of $0{\cdot}025$; calculations were made to six decimal places, the interval δs being $0{\cdot}1$, and the results, given in table 1, are considered accurate to the fifth place, in view of the relatively short range of the integration. Constant checks were made by Weddle's equation, and by finite difference integration of the individual functions $d\phi/ds$, dx/ds and dz/ds over extended intervals by the central difference formula (expressed in terms of backward differences)

$$\int_{x_0}^{x_n} y\,dx = \delta x\left\{\sum_{k=0}^{n} y_k - \tfrac{1}{2}(y_0+y_n) - \tfrac{1}{24}(\Delta y_{n+1}+\Delta y_n-\Delta y_1-\Delta y_0) \right. \\ \left. + \tfrac{11}{1440}(\Delta^3 y_{n+2}+\Delta^3 y_{n+1}-\Delta^3 y_2-\Delta^3 y_1)\ldots\right\}.$$

Further tables were then drawn up, giving x and z in terms of β for each value of the arc distance s, and differenced with respect to β, and the constancy of the higher differences showed that there were no random errors in the fifth decimal place of the calculated figures. These tables were interpolated to derive x and z in terms of s for all other values of β from $-0{\cdot}25$ to $-0{\cdot}6$ at intervals of $0{\cdot}0125$.

For further calculations, the Bessel and Stirling interpolation formulae were used except at the ends of the table, where it was necessary to use the Newton formula. For each value of β, the position of the maximum value of x was derived in terms of s by equating to zero the first differential coefficient obtained from the difference tables. This value was checked at intervals of $0{\cdot}025$ in β by inverse interpolation of the tables of ϕ to give s when $\phi = \frac{1}{2}\pi$. A further check against arithmetical mistakes was made by tabulating these values of s against β and noting the constancy of the higher differences. Values of x_e, the maximum radii, were then calculated by interpolation from the known values of s, and again checked by tabulation against β.

For each value of β, inverse interpolation was employed to find s when $z = 2x_e$, and direct interpolation then gave x_s, the radius at the selected plane. Checks were obtained by tabulating s and x_s against β, and noting the constancy of the higher differences. The values obtained for x_e and x_s are given in table 2, columns 2 and 3.

In each case $S = x_s/x_e$ was calculated, and also $1/H$ by equation (3). The values of S and $1/H$ thus found are given in table 2, columns 4 and 5.

Values of $1/H$ were calculated for S $0{\cdot}66$ $(0{\cdot}01)$ $1{\cdot}00$ by two methods. One method involved inverse interpolation to give the value of β corresponding to each selected value of S, followed by direct interpolation to give the appropriate value of $1/H$. The other method was by interpolation of a table of $1/H$ against S by Newton's formula for unequal intervals, checked periodically by the Lagrange formula to six terms. The former method gave the more reliable results, especially at the ends of the table, but the maximum difference between the values of $1/H$ within the range $S = 0{\cdot}68$ and $S = 0{\cdot}98$ was only 5 in the sixth decimal place. Using values obtained by the former method ordinary subtabulation gave $1/H$ for steps of $0{\cdot}001$ in S, between which linear interpolation is adequate. Table 3 gives the final results thus obtained; for convenience H also has been tabulated against S in table 4.

ACCURACY OF RESULTS

Since difference tables had been drawn up at all stages of the calculation, it was a simple matter to estimate the magnitude of random errors. Six figures had been retained during actual computation and the final tables of S and $1/H$ against β showed that between the limits of 0·68 and 0·98 for S, random errors did not exceed 4 and 3 respectively in the sixth decimal place. The final table of $1/H$ against S showed that the greatest possible random variation in $1/H$ was 7 in the sixth decimal place in the above range. These figures are based on the rather conservative estimate that a random error of ± 1 in a variable will cause random errors not exceeding ± 16 in the sixth difference (the maximum possible error is ± 32).

At the extreme ends of the table, this method of estimation could not be applied. When S exceeded 0·98, in particular, higher differences did not tend rapidly to zero and the accuracy in this region could not be estimated, although there was no reason to suppose that there was any considerable decrease. This was supported by the fact that the different methods used for interpolation did not give variations exceeding 8 in the sixth decimal place. Where S was less than 0·68, accuracy estimates could not be made from the difference tables, but since the higher differences were small, there was no reason to suppose that the errors were greater than in the middle portion of the table. The maximum error found on working back to randomly selected original values of S and $1/H$ was 5 in the sixth decimal place.

The possibility of systematic errors in the calculation was also considered. The highest differences considered in equation (8) were the fifth, and these affected the sixth decimal place of the integrands ϕ, x and z only when β was numerically less than 0·3 and s exceeded 2·5. The largest errors from neglect of the sixth and higher differences would occur with the lowest numerical values of β, and thus of S. Closer consideration of this indicated that the maximum possible error in $1/H$ from this cause was not greater than 1 in the fifth place of decimals when S was less than 0·68, and considerably less elsewhere.

Considering all these possible causes of error, it is considered that the accuracy of the entries in table 3 may be taken to be as follows:

$S = 0{\cdot}66$ to $0{\cdot}68$; maximum error $\pm 0{\cdot}00003 = \pm 0{\cdot}003\,\%$.
$S = 0{\cdot}68$ to $0{\cdot}98$; maximum error $\pm 0{\cdot}00001 = \pm 0{\cdot}001$ to $0{\cdot}003\,\%$.
$S = 0{\cdot}98$ to $1{\cdot}00$; maximum error $\pm 0{\cdot}00003 = \pm 0{\cdot}01\,\%$.

It is concluded that these tables are adequate for all likely experimental purposes, and that the method should be capable of giving results which are as reliable as those obtained by the use of capillary elevations.

It should, however, be pointed out that the attainment of such accuracy would entail very stringent demands upon experimental techniques. For example, if the diameters of the drop are measured to an accuracy of 0·1 %, the error of the calculated surface tension may be almost 0·7 %. For this reason the agreement between the calculated table and the measured values of Andreas *et al.*—the maximum difference is about 1·4 % when $S = 0{\cdot}78$—is considered satisfactory.

Note to tables. In order to avoid possible misreading the headings of the tables have been written in full as s/b, x/b, etc., instead of the dimensionless form used in the text.

TABLE 1. CALCULATED SHAPES OF DROPS

			ϕ deg.			
s/b	ϕ rad.	deg.	min.	sec.	x/b	z/b
			$\beta = -0{\cdot}25$			
0	0		0		0	0
0·1	0·099,97	5	43	40	0·099,83	0·005,00
0·2	0·199,75	11	26	41	0·198,67	0·019,92
0·3	0·299,16	17	8	26	0·295,54	0·044,60
0·4	0·398,01	22	48	16	0·389,48	0·078,75
0·5	0·496,14	28	25	35	0·479,61	0·121,97
0·6	0·593,35	33	59	48	0·565,10	0·173,78
0·7	0·689,50	39	30	20	0·645,19	0·233,60
0·8	0·784,42	44	56	39	0·719,20	0·300,78
0·9	0·877,97	50	18	14	0·786,57	0·374,64
1·0	0·970,00	55	34	37	0·846,80	0·454,42
1·1	1·060,38	60	45	20	0·899,52	0·539,35
1·2	1·148,99	65	49	57	0·944,43	0·628,66
1·3	1·235,71	70	48	2	0·981,35	0·721,56
1·4	1·320,41	75	39	14	1·010,18	0·817,29
1·5	1·402,99	80	23	8	1·030,91	0·915,08
1·6	1·483,34	84	59	21	1·043,62	1·014,25
1·7	1·561,34	89	27	28	1·048,44	1·114,10
1·8	1·636,84	93	46	42	1·045,59	1·214,04
1·9	1·709,72	97	57	34	1·035,34	1·313,49
2·0	1·779,79	101	58	29	1·018,01	1·411,96
2·1	1·846,86	105	49	2	0·993,98	1·509,01
2·2	1·910,65	109	28	21	0·963,64	1·604,28
2·3	1·970,84	112	55	16	0·927,46	1·697,49
2·4	2·027,00	116	8	18	0·885,92	1·788,44
2·5	2·078,53	119	5	28	0·839,53	1·877,02
2·6	2·124,67	121	44	4	0·788,87	1·963,23
2·7	2·164,33	124	0	25	0·734,55	2·047,18
2·8	2·196,01	125	49	20	0·677,26	2·129,13
2·9	2·217,58	127	3	29	0·617,78	2·209,52
3·0	2·225,96	127	32	18	0·557,09	2·289,00
3·1	2·216,67	127	0	22	0·496,39	2·368,47
3·2	2·183,23	125	5	24	0·437,36	2·449,18
			$\beta = -0{\cdot}275$			
0	0		0		0	0
0·1	0·099,97	5	43	40	0·099,83	0·005,00
0·2	0·199,73	11	26	36	0·198,67	0·019,92
0·3	0·299,08	17	8	9	0·295,54	0·044,60
0·4	0·397,82	22	47	36	0·389,49	0·078,73
0·5	0·495,75	28	24	16	0·479,63	0·121,93
0·6	0·592,69	33	57	31	0·565,15	0·173,69
0·7	0·688,46	39	26	45	0·645,29	0·233,44
0·8	0·782,88	44	51	19	0·719,39	0·300,54
0·9	0·875,78	50	10	43	0·786,89	0·374,27
1·0	0·967,03	55	24	24	0·847,33	0·453,89
1·1	1·056,47	60	31	52	0·900,34	0·538,65
1·2	1·143,96	65	32	39	0·945,65	0·627,76
1·3	1·229,38	70	26	18	0·983,09	0·720,45
1·4	1·312,60	75	12	23	1·012,60	0·815,97

TABLE 1 (*cont.*)

		φ deg.				
s/b	φ rad.	deg.	min.	sec.	x/b	z/b
			$\beta = -0{\cdot}275$ (*cont.*)			
1·5	1·393,50	79	50	29	1·034,18	0·913,59
1·6	1·471,95	84	20	12	1·047,92	1·012,61
1·7	1·547,84	88	41	4	1·053,98	1·112,40
1·8	1·621,02	92	52	39	1·052,59	1·212,37
1·9	1·691,34	96	54	23	1·044,05	1·311,98
2·0	1·758,62	100	45	41	1·028,67	1·410,77
2·1	1·822,66	104	25	50	1·006,83	1·508,34
2·2	1·883,19	107	53	56	0·978,97	1·604,37
2·3	1·939,89	111	8	50	0·945,52	1·698,60
2·4	1·992,33	114	9	8	0·906,98	1·790,86
2·5	2·039,98	116	52	57	0·863,87	1·881,08
2·6	2·082,11	119	17	45	0·816,75	1·969,27
2·7	2·117,75	121	20	17	0·766,22	2·055,55
2·8	2·145,62	122	56	5	0·712,97	2·140,19
2·9	2·163,92	123	59	1	0·657,76	2·223,57
3·0	2·170,25	124	20	46	0·601,51	2·306,25
3·1	2·161,28	123	49	55	0·545,34	2·388,98
3·2	2·132,52	122	11	2	0·490,71	2·472,74
			$\beta = -0{\cdot}325$			
0	0		0		0	0
0·1	0·099,96	5	43	39	0·099,83	0·005,00
0·2	0·199,68	11	26	26	0·198,67	0·019,92
0·3	0·298,91	17	7	34	0·295,54	0·044,58
0·4	0·397,42	22	46	14	0·389,50	0·078,69
0·5	0·494,98	28	21	37	0·479,67	0·121,[illegible]
0·6	0·591,37	33	52	58	0·565,24	0·173,52
0·7	0·686,37	39	19	34	0·645,48	0·233,13
0·8	0·779,79	44	40	43	0·719,75	0·300,04
0·9	0·871,43	49	55	45	0·787,53	0·373,52
1·0	0·961,11	55	4	3	0·848,37	0·452,83
1·1	1·048,67	60	5	3	0·901,96	0·537,23
1·2	1·133,95	64	58	13	0·948,06	0·625,93
1·3	1·216,79	69	43	1	0·986,55	0·718,20
1·4	1·297,06	74	18	57	1·017,40	0·813,29
1·5	1·374,62	78	45	35	1·040,65	0·910,52
1·6	1·449,33	83	2	25	1·056,44	1·009,25
1·7	1·521,05	87	8	59	1·064,96	1·108,86
1·8	1·589,63	91	4	45	1·066,48	1·208,83
1·9	1·654,92	94	49	12	1·061,31	1·308,68
2·0	1·716,73	98	21	40	1·049,80	1·408,00
2·1	1·774,84	101	41	27	1·032,36	1·506,45
2·2	1·829,00	104	47	39	1·009,42	1·603,78
2·3	1·878,91	107	39	14	0·981,45	1·699,77
2·4	1·924,20	110	14	54	0·948,94	1·794,33
2·5	1·964,37	112	33	1	0·912,42	1·887,41
2·6	1·998,85	114	31	33	0·872,44	1·979,07
2·7	2·026,89	116	7	57	0·829,60	2·069,43
2·8	2·047,55	117	18	59	0·784,58	2·158,72
2·9	2·059,65	118	0	34	0·738,08	2·247,25
3·0	2·061,71	118	7	38	0·690,95	2·335,44
3·1	2·051,92	117	33	58	0·644,15	2·423,81
3·2	2·028,17	116	12	19	0·598,81	2·512 95

TABLE 1 (*cont.*)

		ϕ deg.				
s/b	ϕ rad.	deg.	min.	sec.	x/b	z/b
			$\beta = -0{\cdot}35$			
0	0		0		0	0
0·1	0·099,96	5	43	38	0·099,83	0·004,99
0·2	0·199,65	11	26	21	0·198,67	0·019,92
0·3	0·298,82	17	7	17	0·295,54	0·044,58
0·4	0·397,22	22	45	33	0·389,51	0·078,68
0·5	0·494,60	28	20	18	0·479,69	0·121,79
0·6	0·590,71	33	50	42	0·565,28	0·173,43
0·7	0·685,33	39	16	0	0·645,57	0·232,97
0·8	0·778,24	44	35	24	0·719,93	0·299,78
0·9	0·869,25	49	48	15	0·787,85	0·373,14
1·0	0·958,15	54	53	53	0·848,89	0·452,30
1·1	1·044,78	59	51	41	0·902,77	0·536,51
1·2	1·128,95	64	41	3	0·949,26	0·625,01
1·3	1·210,52	69	21	27	0·988,27	0·717,06
1·4	1·289,33	73	52	22	1·019,78	0·811,94
1·5	1·365,23	78	13	19	1·043,86	0·908,97
1·6	1·438,09	82	23	46	1·060,66	1·007,52
1·7	1·507,75	86	23	16	1·070,40	1·107,03
1·8	1·574,07	90	11	15	1·073,36	1·206,96
1·9	1·636,88	93	47	11	1·069,87	1·306,89
2·0	1·695,99	97	10	24	1·060,29	1·406,41
2·1	1·751,21	100	20	12	1·045,04	1·505,23
2·2	1·802,26	103	15	43	1·024,55	1·603,10
2·3	1·848,88	105	55	58	0·999,32	1·699,86
2·4	1·890,70	108	19	44	0·969,82	1·795,40
2·5	1·927,28	110	25	30	0·936,60	1·889,72
2·6	1·958,11	112	11	29	0·900,22	1·982,86
2·7	1·982,53	113	35	26	0·861,26	2·074,96
2·8	1·999,76	114	34	40	0·820,40	2·166,23
2·9	2·008,86	115	5	56	0·778,33	2·256,95
3·0	2·008,67	115	5	19	0·735,84	2·347,47
3·1	1·997,92	114	28	20	0·693,84	2·438,22
3·2	1·975,15	113	10	5	0·653,36	2·529,66
			$\beta = -0{\cdot}375$			
0	0		0		0	0
0·1	0·099,95	5	43	37	0·099,83	0·004,99
0·2	0·199,63	11	26	16	0·198,67	0·019,92
0·3	0·298,74	17	7	0	0·295,54	0·044,57
0·4	0·397,02	22	44	52	0·389,51	0·078,66
0·5	0·494,21	28	18	59	0·479,71	0·121,75
0·6	0·590,05	33	48	26	0·565,33	0·173,34
0·7	0·684,29	39	12	25	0·645,67	0·232,82
0·8	0·776,71	44	30	7	0·720,12	0·299,53
0·9	0·867,08	49	40	48	0·788,17	0·372,76
1·0	0·955,21	54	43	46	0·849,41	0·451,77
1·1	1·040,90	59	38	21	0·903,57	0·535,80
1·2	1·123,98	64	23	58	0·950,46	0·624,09
1·3	1·204,27	69	0	0	0·989,98	0·715,92
1·4	1·281,63	73	25	55	1·022,15	0·810,58
1·5	1·355,89	77	41	12	1·047,05	0·907,41
1·6	1·426,91	81	45	21	1·064,87	1·005,79

TABLE 1 (*cont.*)

		ϕ deg.				
s/b	ϕ rad.	deg.	min.	sec.	x/b	z/b
			$\beta = -0{\cdot}375$ (*cont.*)			
1·7	1·494,53	85	37	49	1·075,82	1·105,17
1·8	1·558,60	89	18	4	1·080,21	1·205,05
1·9	1·618,96	92	45	33	1·078,38	1·305,02
2·0	1·675,41	95	59	38	1·070,72	1·404,71
2·1	1·727,76	98	59	35	1·057,64	1·503,84
2·2	1·775,76	101	44	37	1·039,61	1·602,19
2·3	1·819,14	104	13	44	1·017,10	1·699,62
2·4	1·857,56	106	25	50	0·990,63	1·796,05
2·5	1·890,64	108	19	33	0·960,71	1·891,46
2·6	1·917,91	109	53	18	0·927,93	1·985,94
2·7	1·938,82	111	5	11	0·892,88	2·079,59
2·8	1·952,72	111	52	57	0·856,20	2·172,62
2·9	1·958,85	112	14	2	0·818,58	2·265,27
3·0	1·956,37	112	5	30	0·780,79	2·357,86
3·1	1·944,35	111	24	11	0·743,66	2·450,71
3·2	1·921,86	110	6	53	0·708,13	2·544,18
			$\beta = -0{\cdot}425$			
0	0		0		0	0
0·1	0·099,95	5	43	35	0·099,83	0·004,99
0·2	0·199,58	11	26	6	0·198,67	0·019,91
0·3	0·298,57	17	6	25	0·295,55	0·044,56
0·4	0·396,63	22	43	31	0·389,52	0·078,62
0·5	0·493,44	28	16	20	0·479,74	0·121,[illegible]
0·6	0·588,73	33	43	54	0·565,42	0·173,16
0·7	0·682,21	39	5	17	0·645,86	0·232,51
0·8	0·773,64	44	19	34	0·720,48	0·299,03
0·9	0·862,75	49	25	56	0·788,79	0·372,01
1·0	0·949,34	54	23	35	0·850,44	0·450,71
1·1	1·033,18	59	11	48	0·905,17	0·534,37
1·2	1·114,07	63	49	54	0·952,83	0·622,25
1·3	1·191,84	68	17	14	0·993,37	0·713,64
1·4	1·266,31	72	33	14	1·026,85	0·807,84
1·5	1·337,31	76	37	19	1·053,39	0·904,23
1·6	1·404,69	80	28	57	1·073,21	1·002,23
1·7	1·468,28	84	7	33	1·086,56	1·101,32
1·8	1·527,92	87	32	36	1·093,79	1·201,04
1·9	1·583,45	90	43	29	1·095,27	1·301,02
2·0	1·634,68	93	39	36	1·091,41	1·400,93
2·1	1·681,42	96	20	18	1·082,66	1·500,54
2·2	1·723,45	98	44	48	1·069,49	1·599,66
2·3	1·760,53	100	52	16	1·052,41	1·698,19
2·4	1·792,38	102	41	45	1·031,95	1·796,07
2·5	1·818,69	104	12	12	1·008,64	1·893,31
2·6	1·839,10	105	22	22	0·983,07	1·989,98
2·7	1·853,24	106	10	58	0·955,83	2·086,20
2·8	1·860,68	106	36	34	0·927,55	2·182,12
2·9	1·860,99	106	37	36	0·898,89	2·277,92
3·0	1·853,72	106	12	38	0·870,56	2·373,82
3·1	1·838,49	105	20	16	0·843,31	2·470,04
3·2	1·814,98	103	59	26	0·817,93	2·566,76

TABLE 1 (*cont.*)

		ϕ deg.				
s/b	ϕ rad.	deg.	min.	sec.	x/b	z/b
			$\beta = -0·45$			
0	0		0		0	0
0·1	0·099,94	5	43	35	0·099,83	0·004,99
0·2	0·199,55	11	26	1	0·198,67	0·019,91
0·3	0·298,49	17	6	8	0·295,55	0·044,55
0·4	0·396,43	22	42	50	0·389,53	0·078,60
0·5	0·493,06	28	15	1	0·479,76	0·121,62
0·6	0·588,07	33	41	38	0·565,46	0·173,07
0·7	0·681,18	39	1	42	0·645,95	0·232,35
0·8	0·772,10	44	14	17	0·720,66	0·298,78
0·9	0·860,59	49	18	29	0·789,11	0·371,64
1·0	0·946,40	54	13	29	0·850,96	0·450,18
1·1	1·029,32	58	58	32	0·905,97	0·533,65
1·2	1·109,13	63	32	54	0·954,01	0·621,32
1·3	1·185,64	67	55	56	0·995,06	0·712,48
1·4	1·258,68	72	7	1	1·029,19	0·806,45
1·5	1·328,07	76	5	34	1·056,54	0·902,62
1·6	1·393,64	79	50	59	1·077,35	1·000,41
1·7	1·455,24	83	22	45	1·091,89	1·099,33
1·8	1·512,70	86	40	17	1·100,53	1·198,94
1·9	1·565,86	89	43	1	1·103,64	1·298,88
2·0	1·614,53	92	30	20	1·101,67	1·398,85
2·1	1·658,52	95	1	35	1·095,06	1·498,63
2·2	1·697,64	97	16	4	1·084,31	1·598,04
2·3	1·731,65	99	12	59	1·069,93	1·697,00
2·4	1·760,31	100	51	30	1·052,46	1·795,46
2·5	1·783,34	102	10	40	1·032,44	1·893,43
2·6	1·800,44	103	9	28	1·010,47	1·990,99
2·7	1·811,31	103	46	49	0·987,13	2·088,22
2·8	1·815,62	104	1	38	0·963,04	2·185,28
2·9	1·813,05	103	52	48	0·938,87	2·282,31
3·0	1·803,32	103	19	21	0·915,30	2·379,49
3·1	1·786,21	102	20	32	0·893,03	2·476,98
3·2	1·761,59	100	55	55	0·872,79	2·574,91
			$\beta = -0·475$			
0	0		0		0	0
0·1	0·099,94	5	43	34	0·099,83	0·004,99
0·2	0·199,53	11	25	55	0·198,67	0·019,91
0·3	0·298,40	17	5	50	0·295,55	0·044,55
0·4	0·396,23	22	42	9	0·389,54	0·078,58
0·5	0·492,67	28	13	41	0·479,78	0·121,57
0·6	0·587,41	33	39	22	0·565,51	0·172,98
0·7	0·680,14	38	58	9	0·646,05	0·232,20
0·8	0·770,57	44	9	1	0·720,84	0·298,53
0·9	0·858,43	49	11	4	0·789,42	0·371,26
1·0	0·943,48	54	3	26	0·851,47	0·449,64
1·1	1·025,47	58	45	19	0·906,76	0·532,93
1·2	1·104,21	63	15	59	0·955,19	0·620,39
1·3	1·179,47	67	34	43	0·996,74	0·711,32
1·4	1·251,08	71	40	54	1·031,52	0·805,06
1·5	1·318,87	75	33	56	1·059,67	0·901,00
1·6	1·382,66	79	13	13	1·081,46	0·998,58

TABLE 1 (*cont.*)

		ϕ deg.				
s/b	ϕ rad.	deg.	min.	sec.	x/b	z/b
			$\beta = -0{\cdot}475$ (*cont.*)			
1·7	1·442,28	82	38	12	1·097,19	1·097,32
1·8	1·497,58	85	48	17	1·107,22	1·196,80
1·9	1·548,38	88	42	57	1·111,96	1·296,67
2·0	1·594,52	91	21	33	1·111,86	1·396,67
2·1	1·635,81	93	43	30	1·107,38	1·496,56
2·2	1·672,05	95	48	6	1·099,04	1·596,20
2·3	1·703,05	97	34	39	1·087,35	1·695,51
2·4	1·728,57	99	2	24	1·072,85	1·794,45
2·5	1·748,39	100	10	32	1·056,11	1·893,04
2·6	1·762,26	100	58	12	1·037,71	1·991,33
2·7	1·769,92	101	24	33	1·018,26	2·089,42
2·8	1·771,15	101	28	46	0·998,36	2·187,42
2·9	1·765,73	101	10	8	0·978,67	2·285,47
3·0	1·753,50	100	28	6	0·959,84	2·383,68
3·1	1·734,37	99	22	18	0·942,56	2·482,17
3·2	1·708,37	97	52	57	0·927,50	2·581,03
			$\beta = -0{\cdot}525$			
0	0		0		0	0
0·1	0·099,93	5	43	33	0·099,83	0·004,99
0·2	0·199,48	11	25	45	0·198,67	0·019,91
0·3	0·298,24	17	5	16	0·295,55	0·044,54
0·4	0·395,84	22	40	47	0·389,55	0·078,54
0·5	0·491,91	28	11	3	0·479,82	0·121,48
0·6	0·586,10	33	34	51	0·565,60	0·172,81
0·7	0·678,07	38	51	2	0·646,24	0·231,88
0·8	0·767,51	43	58	31	0·721,20	0·298,02
0·9	0·854,13	48	56	17	0·790,05	0·370,51
1·0	0·937,65	53	43	24	0·852,49	0·448,58
1·1	1·017,82	58	18	59	0·908,34	0·531,49
1·2	1·094,40	62	42	17	0·957,53	0·618,53
1·3	1·167,19	66	52	29	1·000,09	0·709,00
1·4	1·235,98	70	48	59	1·036,14	0·802,25
1·5	1·300,58	74	31	5	1·065,89	0·897,71
1·6	1·360,83	77	58	12	1·089,63	0·994,83
1·7	1·416,55	81	9	45	1·107,70	1·093,18
1·8	1·467,58	84	5	11	1·120,49	1·192,34
1·9	1·513,76	86	43	56	1·128,46	1·292,02
2·0	1·554,92	89	5	26	1·132,06	1·391,95
2·1	1·590,91	91	9	8	1·131,80	1·491,94
2·2	1·621,54	92	54	27	1·128,21	1·591,87
2·3	1·646,66	94	20	48	1·121,84	1·691,67
2·4	1·666,09	95	27	37	1·113,25	1·791,29
2·5	1·679,68	96	14	20	1·103,00	1·890,77
2·6	1·687,27	96	40	24	1·091,71	1·990,13
2·7	1·688,72	96	45	25	1·079,97	2·089,43
2·8	1·683,95	96	28	59	1·068,39	2·188,76
2·9	1·672,90	95	51	0	1·057,59	2·288,18
3·0	1·655,58	94	51	29	1·048,21	2·387,73
3·1	1·632,12	93	30	48	1·040,86	2·487,46
3·2	1·602,70	91	49	40	1·036,15	2·587.35

TABLE 1 (*cont.*)

			ϕ deg.			
s/b	ϕ rad.	deg.	min.	sec.	x/b	z/b
			$\beta = -0{\cdot}55$			
0	0		0		0	0
0·1	0·099,93	5	43	33	0·099,83	0·004,99
0·2	0·199,45	11	25	40	0·198,67	0·019,91
0·3	0·298,15	17	4	58	0·295,55	0·044,53
0·4	0·395,64	22	40	6	0·389,56	0·078,53
0·5	0·491,52	28	9	44	0·479,84	0·121,44
0·6	0·585,44	33	32	36	0·565,65	0·172,72
0·7	0·677,04	38	47	29	0·646,34	0·231,73
0·8	0·765,99	43	53	16	0·721,38	0·297,78
0·9	0·851,98	48	48	54	0·790,36	0·370,13
1·0	0·934,74	53	33	24	0·853,00	0·448,05
1·1	1·014,00	58	5	52	0·909,13	0·530,78
1·2	1·089,52	62	25	29	0·958,69	0·617,60
1·3	1·161,07	66	31	28	1·001,75	0·707,84
1·4	1·228,46	70	23	9	1·038,43	0·800,85
1·5	1·291,50	73	59	50	1·068,97	0·896,05
1·6	1·349,99	77	20	56	1·093,67	0·992,94
1·7	1·403,79	80	25	52	1·112,90	1·091,06
1·8	1·452,71	83	14	4	1·127,07	1·190,04
1·9	1·496,62	85	44	59	1·136,62	1·289,58
2·0	1·535,34	87	58	6	1·142,06	1·389,42
2·1	1·568,72	89	52	51	1·143,89	1·489,40
2·2	1·596,61	91	28	45	1·142,66	1·589,39
2·3	1·618,86	92	45	15	1·138,91	1·689,32
2·4	1·635,34	93	41	52	1·133,24	1·789,16
2·5	1·645,90	94	18	11	1·126,21	1·888,91
2·6	1·650,43	94	33	47	1·118,43	1·988,61
2·7	1·648,87	94	28	24	1·110,50	2·088,29
2·8	1·641,17	94	1	55	1·103,04	2·188,01
2·9	1·627,34	93	14	22	1·096,64	2·287,80
3·0	1·607,47	92	6	3	1·091,94	2·387,69
3·1	1·581,72	90	37	34	1·089,51	2·487,66
3·2	1·550,37	88	49	46	1·089,94	2·587,65
			$\beta = -0{\cdot}575$			
0	0		0		0	0
0·1	0·099,93	5	43	32	0·099,83	0·004,99
0·2	0·199,43	11	25	35	0·198,67	0·019,91
0·3	0·298,07	17	4	41	0·295,56	0·044,52
0·4	0·395,44	22	39	25	0·389,56	0·078,51
0·5	0·491,14	28	8	25	0·479,85	0·121,40
0·6	0·584,79	33	30	20	0·565,69	0·172,63
0·7	0·676,01	38	43	56	0·646,43	0·231,57
0·8	0·764,46	43	48	2	0·721,55	0·297,53
0·9	0·849,84	48	41	32	0·790,67	0·369,76
1·0	0·931,84	53	23	25	0·853,50	0·447,51
1·1	1·010,19	57	52	46	0·909,91	0·530,06
1·2	1·084,64	62	8	44	0·959,85	0·616,67
1·3	1·154,98	66	10	31	1·003,40	0·706,66
1·4	1·220,97	69	57	23	1·040,71	0·799,42
1·5	1·282,44	73	28	42	1·072,04	0·894,37
1·6	1·339,20	76	43	50	1·097,70	0·991,01

TABLE 1 (*cont.*)

s/b	ϕ rad.	ϕ deg. deg.	min.	sec.	x/b	z/b
		$\beta = -0{\cdot}575$ (*cont.*)				
1·7	1·391,09	79	42	12	1·118,08	1·088,90
1·8	1·437,93	82	23	14	1·133,60	1·187,68
1·9	1·479,58	84	46	25	1·144,74	1·287,05
2·0	1·515,89	86	51	14	1·151,99	1·386,78
2·1	1·546,70	88	37	10	1·155,89	1·486,70
2·2	1·571,89	90	3	46	1·157,00	1·586,69
2·3	1·591,33	91	10	36	1·155,87	1·686,68
2·4	1·604,90	91	57	14	1·153,09	1·786,64
2·5	1·612,50	92	23	21	1·149,25	1·886,57
2·6	1·614,05	92	28	42	1·144,95	1·986,47
2·7	1·609,53	92	13	10	1·140,80	2·086,39
2·8	1·598,95	91	36	46	1·137,41	2·186,33
2·9	1·582,37	90	39	46	1·135,37	2·286,31
3·0	1·559,94	89	22	40	1·135,29	2·386,30
3·1	1·531,88	87	46	12	1·137,73	2·486,27
3·2	1·498,48	85	51	23	1·143,25	2·586,11

TABLE 2. CALCULATED FACTORS FOR PENDANT DROPS

$-\beta$	x_s/b	x_e/b	S	$1/H$
0·2500	1·048,50	0·700,18	0·667,81	0·909,63
0·2625	1·051,39	0·716,78	0·681,74	0·861,56
0·2750	1·054,33	0·733,18	0·695,40	0·817,81
0·2875	1·057,34	0·749,44	0·708,80	0·777,8[illegible]
0·3000	1·060,41	0·765,56	0·721,95	0·741,10
0·3125	1·063,54	0·781,59	0·734,89	0·707,27
0·3250	1·066,74	0·797,53	0·747,63	0·675,99
0·3375	1·070,02	0·813,42	0·760,19	0·646,97
0·3500	1·073,37	0·829,28	0·772,59	0·619,97
0·3625	1·076,81	0·845,13	0·784,84	0·594,78
0·3750	1·080,33	0·860,99	0·796,97	0·571,21
0·3875	1·083,95	0·876,89	0·808,98	0·549,10
0·4000	1·087,66	0·892,82	0·820,86	0·528,31
0·4125	1·091,49	0·908,83	0·832,66	0·508,72
0·4250	1·095,42	0·924,94	0·844,37	0·490,22
0·4375	1·099,48	0·941,17	0·856,01	0·472,71
0·4500	1·103,67	0·957,53	0·867,59	0·456,09
0·4625	1·108,00	0·974,05	0·879,11	0·440,30
0·4750	1·112,49	0·990,76	0·890,57	0·425,26
0·4875	1·117,15	1·007,68	0·902,01	0·410,90
0·5000	1·122,01	1·024,84	0·913,40	0·397,17
0·5125	1·127,08	1·042,28	0·924,76	0·384,01
0·5250	1·132,39	1·060,03	0·936,10	0·371,35
0·5375	1·137,98	1·078,13	0·947,41	0·359,16
0·5500	1·143,90	1·096,65	0·958,69	0·347,38
0·5625	1·150,20	1·115,62	0·969,93	0·335,95
0·5750	1·157,00	1·135,13	0·981,10	0·324,79
0·5875	1·164,44	1·155,28	0·992,13	0·313,83
0·6000	1·172,87	1·176,23	1·002,87	0·302,89

TABLE 3. TABLE OF $1/H$ IN TERMS OF S

S	0	1	2	3	4	5	6	7	8	9
0·66	0·938,28	0·934,54	0·930,82	0·927,12	0·923,45	0·919,79	0·916,16	0·912,55	0·908,95	0·905,38
0·67	0·901,83	0·898,30	0·894,78	0·891,29	0·887,82	0·884,36	0·880,92	0·877,51	0·874,11	0·870,73
0·68	0·867,37	0·864,03	0·860,70	0·857,39	0·854,10	0·850,83	0·847,58	0·844,34	0·841,12	0·837,92
0·69	0·834,73	0·831,56	0·828,41	0·825,27	0·822,15	0·819,05	0·815,96	0·812,89	0·809,83	0·806,79
0·70	0·803,76	0·800,75	0·797,76	0·794,78	0·791,82	0·788,87	0·785,94	0·783,02	0·780,11	0·777,22
0·71	0·774,35	0·771,49	0·768,64	0·765,81	0·763,00	0·760,19	0·757,41	0·754,63	0·751,87	0·749,12
0·72	0·746,39	0·743,67	0·740,97	0·738,28	0·735,60	0·732,93	0·730,28	0·727,64	0·725,02	0·722,40
0·73	0·719,80	0·717,22	0·714,64	0·712,08	0·709,53	0·707,00	0·704,47	0·701,96	0·699,46	0·696,97
0·74	0·694,49	0·692,02	0·689,57	0·687,13	0·684,70	0·682,28	0·679,88	0·677,48	0·675,10	0·672,73
0·75	0·670,37	0·668,03	0·665,69	0·663,37	0·661,05	0·658,75	0·656,46	0·654,18	0·651,91	0·649,65
0·76	0·647,40	0·645,16	0·642,94	0·640,72	0·638,51	0·636,32	0·634,13	0·631,95	0·629,79	0·627,63
0·77	0·625,49	0·623,35	0·621,22	0·619,11	0·617,00	0·614,90	0·612,81	0·610,74	0·608,67	0·606,61
0·78	0·604,57	0·602,53	0·600,50	0·598,48	0·596,47	0·594,47	0·592,48	0·590,49	0·588,52	0·586,56
0·79	0·584,60	0·582,65	0·580,72	0·578,79	0·576,87	0·574,96	0·573,05	0·571,16	0·569,27	0·567,39
0·80	0·565,53	0·563,66	0·561,81	0·559,97	0·558,13	0·556,30	0·554,48	0·552,66	0·550,86	0·549,06
0·81	0·547,27	0·545,49	0·543,71	0·541,95	0·540,19	0·538,44	0·536,69	0·534,96	0·533,23	0·531,51
0·82	0·529,79	0·528,08	0·526,38	0·524,69	0·523,00	0·521,32	0·519,65	0·517,99	0·516,34	0·514,69
0·83	0·513,05	0·511,42	0·509,79	0·508,17	0·506,56	0·504,96	0·503,36	0·501,76	0·500,18	0·498,60
0·84	0·497,03	0·495,46	0·493,90	0·492,34	0·490,80	0·489,26	0·487,72	0·486,19	0·484,67	0·483,16
0·85	0·481,65	0·480,15	0·478,65	0·477,16	0·475,67	0·474,20	0·472,72	0·471,26	0·469,80	0·468,34
0·86	0·466,90	0·465,45	0·464,02	0·462,59	0·461,16	0·459,74	0·458,33	0·456,92	0·455,52	0·454,12
0·87	0·452,73	0·451,34	0·449,96	0·448,58	0·447,21	0·445,84	0·444,48	0·443,13	0·441,78	0·440,44
0·88	0·439,10	0·437,77	0·436,44	0·435,12	0·433,80	0·432,49	0·431,18	0·429,88	0·428,58	0·427,29
0·89	0·426,00	0·424,71	0·423,44	0·422,16	0·420,89	0·419,63	0·418,37	0·417,12	0·415,87	0·414,62
0·90	0·413,38	0·412,14	0·410,91	0·409,68	0·408,46	0·407,24	0·406,02	0·404,81	0·403,60	0·402,40
0·91	0·401,21	0·400,01	0·398,82	0·397,64	0·396,46	0·395,28	0·394,11	0·392,94	0·391,77	0·390,61
0·92	0·389,46	0·388,31	0·387,16	0·386,01	0·384,87	0·383,74	0·382,60	0·381,47	0·380,35	0·379,22
0·93	0·378,10	0·376,99	0·375,88	0·374,77	0·373,66	0·372,56	0·371,46	0·370,37	0·369,28	0·368,19
0·94	0·367,11	0·366,02	0·364,94	0·363,87	0·362,80	0·361,73	0·360,66	0·359,60	0·358,54	0·357,48
0·95	0·356,43	0·355,38	0·354,33	0·353,28	0·352,24	0·351,20	0·350,16	0·349,13	0·348,09	0·347,06
0·96	0·346,04	0·345,01	0·343,99	0·342,97	0·341,95	0·340,93	0·339,92	0·338,90	0·337,89	0·336,88
0·97	0·335,88	0·334,87	0·333,87	0·332,87	0·331,86	0·330,86	0·329,87	0·328,87	0·327,87	0·326,88
0·98	0·325,88	0·324,89	0·323,89	0·322,90	0·321,91	0·320,92	0·319,92	0·318,93	0·317,94	0·316,95
0·99	0·315,95	0·314,96	0·313,96	0·312,96	0·311,96	0·310,95	0·309,94	0·308,93	0·307,92	0·306,90
1·00	0·305,88	0·304,84	0·303,81	0·302,76	—	—	—	—	—	—

TABLE 4. TABLE OF H IN TERMS OF S

S	0	1	2	3	4	5	6	7	8	9
0·66	1·065,79	1·070,05	1·074,32	1·078,61	1·082,90	1·087,20	1·091,51	1·095,84	1·100,17	1·104,51
0·67	1·108,86	1·113,22	1·117,59	1·121,97	1·126,36	1·130,76	1·135,17	1·139,59	1·144,02	1·148,46
0·68	1·152,91	1·157,37	1·161,85	1·166,33	1·170,82	1·175,32	1·179,83	1·184,36	1·188,89	1·193,44
0·69	1·197,99	1·202,56	1·207,13	1·211,72	1·216,32	1·220,93	1·225,55	1·230,18	1·234,83	1·239,48
0·70	1·244,15	1·248,82	1·253,51	1·258,21	1·262,92	1·267,64	1·272,37	1·277,11	1·281,86	1·286,63
0·71	1·291,41	1·296,19	1·300,99	1·305,80	1·310,62	1·315,46	1·320,30	1·325,15	1·330,02	1·334,89
0·72	1·339,78	1·344,68	1·349,59	1·354,51	1·359,44	1·364,38	1·369,33	1·374,30	1·379,28	1·384,27
0·73	1·389,27	1·394,28	1·399,30	1·404,33	1·409,38	1·414,44	1·419,51	1·424,59	1·429,68	1·434,79
0·74	1·439,91	1·445,04	1·450,18	1·455,33	1·460,49	1·465,67	1·470,85	1·476,05	1·481,26	1·486,48
0·75	1·491,7	1·496,9	1·502,2	1·507,5	1·512,7	1·518,0	1·523,3	1·528,6	1·534,0	1·539,3
0·76	1·544,6	1·550,0	1·555,4	1·560,7	1·566,1	1·571,5	1·577,0	1·582,4	1·587,8	1·593,3
0·77	1·598,8	1·604,2	1·609,7	1·615,2	1·620,8	1·626,3	1·631,8	1·637,4	1·642,9	1·648,5
0·78	1·654,1	1·659,7	1·665,3	1·670,9	1·676,5	1·682,2	1·687,8	1·693,5	1·699,2	1·704,9
0·79	1·710,6	1·716,3	1·722,0	1·727,7	1·733,5	1·739,3	1·745,0	1·750,8	1·756,6	1·762,4
0·80	1·768,3	1·774,1	1·780,0	1·785,8	1·791,7	1·797,6	1·803,5	1·809,4	1·815,3	1·821,3
0·81	1·827,3	1·833,2	1·839,2	1·845,2	1·851,2	1·857,2	1·863,3	1·869,3	1·875,4	1·881,4
0·82	1·887,6	1·893,7	1·899,8	1·905,9	1·912,0	1·918,2	1·924,4	1·930,5	1·936,7	1·942,9
0·83	1·949,1	1·955,3	1·961,6	1·967,8	1·974,1	1·980,4	1·986,7	1·993,0	1·999,3	2·005,6
0·84	2·012,0	2·018,3	2·024,7	2·031,1	2·037,5	2·043,9	2·050,3	2·056,8	2·063,3	2·069,7
0·85	2·076,2	2·082,7	2·089,2	2·095,7	2·102,3	2·108,8	2·115,4	2·122,0	2·128,6	2·135,2
0·86	2·141,8	2·148,4	2·155,1	2·161,8	2·168,4	2·175,1	2·181,8	2·188,6	2·195,3	2·202,1
0·87	2·208,8	2·215,6	2·222,4	2·229,3	2·236,1	2·242,9	2·249,8	2·256,7	2·263,6	2·270,5
0·88	2·277,4	2·284,3	2·291,3	2·298,2	2·305,2	2·312,2	2·319,2	2·326,3	2·333,3	2·340,4
0·89	2·347,4	2·354,5	2·361,6	2·368,8	2·375,9	2·383,1	2·390,2	2·397,4	2·404,6	2·411,9
0·90	2·419,1	2·426,4	2·433,6	2·440,9	2·448,3	2·455,6	2·462,9	2·470,3	2·477,7	2·485,1
0·91	2·492,5	2·499,9	2·507,4	2·514,9	2·522,3	2·529,9	2·537,4	2·544,9	2·552,5	2·560,1
0·92	2·567,7	2·575,3	2·582,9	2·590,6	2·598,3	2·606,0	2·613,7	2·621,4	2·629,2	2·637,0
0·93	2·644,8	2·652,6	2·660,4	2·668,3	2·676,2	2·684,1	2·692,1	2·700,0	2·708,0	2·716,0
0·94	2·724,0	2·732,1	2·740,1	2·748,2	2·756,4	2·764,5	2·772,7	2·780,9	2·789,1	2·797,3
0·95	2·805,6	2·813,9	2·822,2	2·830,6	2·839,0	2·847,4	2·855,8	2·864,3	2·872,8	2·881,3
0·96	2·889,9	2·898,5	2·907,1	2·915,7	2·924,4	2·933,1	2·941,9	2·950,7	2·959,5	2·968,4
0·97	2·977,3	2·986,2	2·995,2	3·004,2	3·013,3	3·022,4	3·031,5	3·040,7	3·050,0	3·059,3
0·98	3·068,6	3·078,0	3·087,4	3·096,9	3·106,5	3·116,1	3·125,7	3·135,5	3·145,2	3·155,1
0·99	3·165,0	3·175,0	3·185,1	3·195,3	3·205,6	3·215,9	3·226,4	3·236,9	3·247,6	3·258,4
1·00	3·269,3	3·280,4	3·291,6	3·302,9	—	—	—	—	—	—

REFERENCES

Adam, N. K. 1938 *The physics and chemistry of surfaces*, 2nd ed., pp. 363–388. Oxford: Clarendon Press.

Andreas, J. M., Hauser, E. A. & Tucker, W. B. 1938 *J. Phys. Chem.* **42**, 1001.

Bashforth, F. & Adams, J. C. 1883 *An attempt to test the theories of capillary action.* Cambridge University Press.

De Noüy, P. L. 1919 *J. Gen. Physiol.* **1**, 521.

Ferguson, A. 1912 *Phil. Mag.* **23**, 417.

Harkins, W. D. & Brown, F. E. 1919 *J. Amer. Chem. Soc.* **41**, 499.

Sudgen, S. 1921 *J. Chem. Soc.* p. 1483.

Sugden, S. 1922 *J. Chem. Soc.* p. 858.

Wilhelmy, L. 1863 *Ann. Phys., Lpz.*, **9**, 475.

Worthington, A. M. 1881 *Proc. Roy. Soc.* **32**, 362.

Worthington, A. M. 1885 *Phil. Mag.* **19**, 46.

Concerning the effect of compressibility on laminar boundary layers and their separation

By L. HOWARTH, *St John's College, Cambridge*

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 20 *November* 1947)

The theory of compressible flow in a laminar boundary layer has been developed for the case when the viscosity is assumed to be proportional to the absolute temperature and the Prandtl number is unity. (These assumptions may be compared with the empirical relations $\mu \propto T^{\frac{8}{9}}$ and $\sigma = 0{\cdot}715$ suggested by Cope.) It is shown that a transformation of the ordinate normal to the layer can lead to a simplified form of equation of motion very similar to the ordinary incompressible equation but modified by a multiplicative factor G in the pressure term. This factor is greater than unity at the boundary and tends to one at the outside of the layer.

Several particular solutions are considered including accelerated flow with a linearly increasing velocity and retarded flow along a flat plate with a linearly decreasing velocity.

The general implications of the theory are discussed and qualitative conclusions are drawn when the mainstream velocity starts from a stagnation point, rises to a maximum and subsequently falls. It is concluded that for such a velocity distribution increasing compressibility will reduce the skin friction, increase the boundary layer thickness and cause earlier separation as compared with the incompressible flow with the same mainstream velocity distribution and the kinematic viscosity corresponding to conditions at the stagnation point.

1. INTRODUCTION

The recent work of Cope & Hartree (1948) has made it abundantly clear that a complete study of compressible flow in boundary layers when allowance is made for the empirical temperature variation of viscosity and conductivity is a matter for modern electronic calculating machines. These authors have, in fact, initiated a study of the flow along a flat plate in the presence of a linear retarding pressure gradient by this means.

The empirical relations chosen by Cope (unpublished) for air in the temperature range 90° K to 300° K (the range important in wind tunnel experiments) are

$$\mu \propto T^{\frac{8}{9}} \quad \text{and} \quad \sigma = 0{\cdot}715, \tag{1}$$

where μ is the viscosity, T is the absolute temperature and σ is the Prandtl number defined as $\mu c_p/k$ where k is the thermal conductivity and c_p is the specific heat at constant pressure.

The difficulty in any numerical approach to a problem of this complexity lies in the number of particular solutions required to give a full understanding of the effects of compressibility on boundary layers. It was therefore thought to be worthwhile to develop the theory when the empirical relations are replaced by

$$\mu \propto T, \tag{2}$$

and

$$\sigma = 1, \tag{3}$$

for it appears that in this case considerable simplifications in the theory can be achieved. It is thought that these assumptions are sufficiently close to Cope's empirical relations to make the results at least of qualitative interest.

The assumption $\sigma = 1$ has been made by many previous writers and leads, even in the case of variable viscosity, to stagnation temperature at a heat insulating boundary (Crocco 1946). The assumption $\mu \propto T$ has effectively been made by Crocco for the flow along a flat plate in the absence of a pressure gradient when he assumed $\rho\mu$ constant. The effects of these two assumptions taken together appears, in view of the stagnation temperature at the boundary, to lead to an overestimate of viscous effects in the boundary layer.

Accepting these two basic assumptions the theory will be developed exactly and with particular reference to the effects of pressure gradients on boundary layers along heat insulating boundaries.

2. THE EQUATIONS OF MOTION AND CONTINUITY

The equations of motion and continuity are

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial x}+\frac{1}{\rho}\frac{\partial}{\partial y}\left(\mu\frac{\partial \mu}{\partial y}\right), \tag{4}$$

$$0 = \frac{\partial p}{\partial y}, \tag{5}$$

$$\frac{\partial}{\partial x}(\rho u)+\frac{\partial}{\partial y}(\rho v) = 0. \tag{6}$$

Equation (6) implies that a stream function ψ exists such that

$$\rho u = \rho_s\frac{\partial \psi}{\partial y}, \quad \rho v = -\rho_s\frac{\partial \psi}{\partial x}, \tag{7}$$

where the suffix s is used to denote some standard condition of the gas, at some specified point of the boundary layer or main flow outside. The viscosity will also be defined in terms of the temperature T_s at this point by the relation

$$\mu = \mu_s\frac{T}{T_s}. \tag{8}$$

Let us now change from independent variables x, y to x, Y where

$$Y = \int_0^y \left(\frac{\nu_s}{\nu}\right)^{\frac{1}{2}} dy = \left(\frac{p}{p_s}\right)^{\frac{1}{2}} \int_0^y \frac{T_s}{T} dy \tag{9}$$

and p, T denote the pressure and temperature at any point inside the boundary layer. The pressure p being independent of y by virtue of equation (5) has been taken outside the integral sign. Furthermore let us modify ψ by writing

$$\psi(x, y) = \left(\frac{p}{p_s}\right)^{\frac{1}{2}} \chi(x, Y). \tag{10}$$

Then

$$u = \frac{\rho_s}{\rho}\left(\frac{\partial \psi}{\partial y}\right)_x = \frac{\rho_s}{\rho}\left(\frac{p}{p_s}\right)^{\frac{1}{2}} \left(\frac{\partial \chi}{\partial Y}\right)_x \left(\frac{\partial Y}{\partial y}\right)_x, \tag{11}$$

where subscripts x, y, Y are used to show which quantities remain constant in a partial differentiation. Hence

$$u = \frac{\rho_s}{\rho}\left(\frac{p}{p_s}\right)^{\frac{1}{2}} \frac{\partial \chi}{\partial Y}\left(\frac{p}{p_s}\right)^{\frac{1}{2}} \frac{T_s}{T} = \frac{p\rho_s T_s}{p_s \rho T}\frac{\partial \chi}{\partial Y} = \frac{\partial \chi}{\partial Y} \tag{12}$$

assuming the gas to be perfect. Next

$$v = -\frac{\rho_s}{\rho}\left(\frac{\partial \psi}{\partial x}\right)_y = -\frac{1}{2}\frac{\rho_s}{p_s \rho}\left(\frac{p_s}{p}\right)^{\frac{1}{2}} \frac{\partial p}{\partial x}\chi - \left(\frac{p}{p_s}\right)^{\frac{1}{2}} \frac{\rho_s}{\rho}\left[\left(\frac{\partial \chi}{\partial x}\right)_Y + \left(\frac{\partial \chi}{\partial Y}\right)_x \left(\frac{\partial Y}{\partial x}\right)_y\right], \tag{13}$$

$$\left(\frac{\partial u}{\partial x}\right)_y = \left(\frac{\partial u}{\partial x}\right)_Y + \left(\frac{\partial u}{\partial Y}\right)_x \left(\frac{\partial Y}{\partial x}\right)_y = \frac{\partial^2 \chi}{\partial x \partial Y} + \frac{\partial^2 \chi}{\partial Y^2}\left(\frac{\partial Y}{\partial x}\right)_y, \tag{14}$$

$$\left(\frac{\partial u}{\partial y}\right)_x = \left(\frac{\partial u}{\partial Y}\right)_x \left(\frac{\partial Y}{\partial y}\right)_x = \frac{\partial^2 \chi}{\partial Y^2}\left(\frac{\partial Y}{\partial y}\right)_x = \left(\frac{p}{p_s}\right)^{\frac{1}{2}} \frac{T_s}{T}\frac{\partial^2 \chi}{\partial Y^2}. \tag{15}$$

Hence

$$\begin{aligned} u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} &= \frac{\partial \chi}{\partial Y}\frac{\partial^2 \chi}{\partial x \partial Y} + \frac{\partial \chi}{\partial Y}\frac{\partial^2 \chi}{\partial Y^2}\left(\frac{\partial Y}{\partial x}\right)_y - \frac{1}{2p_s}\frac{\rho_s T_s}{\rho T}\chi\frac{\partial^2 \chi}{\partial Y^2}\frac{\partial p}{\partial x} \\ &\quad - \frac{p\rho_s T_s}{p_s \rho T}\left[\frac{\partial \chi}{\partial x}\frac{\partial^2 \chi}{\partial Y^2} + \frac{\partial \chi}{\partial Y}\frac{\partial^2 \chi}{\partial Y^2}\left(\frac{\partial Y}{\partial x}\right)_y\right] \\ &= \frac{\partial^2 \chi}{\partial x \partial Y}\frac{\partial \chi}{\partial Y} - \frac{\partial^2 \chi}{\partial Y^2}\frac{\partial \chi}{\partial x} - \frac{1}{2p}\frac{\partial p}{\partial x}\chi\frac{\partial^2 \chi}{\partial Y^2} \end{aligned} \tag{16}$$

again making use of the perfect gas law.

Now from equation (15)

$$\mu\left(\frac{\partial u}{\partial y}\right)_x = \frac{\mu_s T}{T_s}\left(\frac{\partial u}{\partial y}\right)_x = \mu_s\left(\frac{p}{p_s}\right)^{\frac{1}{2}} \frac{\partial^2 \chi}{\partial Y^2}, \tag{17}$$

so that

$$\begin{aligned} \frac{1}{\rho}\left[\frac{\partial}{\partial y}\left(\mu\frac{\partial u}{\partial y}\right)\right]_x &= \frac{T_s}{\rho T}\left(\frac{p}{p_s}\right)^{\frac{1}{2}} \frac{\partial}{\partial Y}\left\{\mu_s\left(\frac{p}{p_s}\right)^{\frac{1}{2}} \frac{\partial^2 \chi}{\partial Y^2}\right\} \\ &= \frac{\mu_s p T_s}{\rho p_s T}\frac{\partial^3 \chi}{\partial Y^3} = \nu_s\frac{\partial^3 \chi}{\partial Y^3}. \end{aligned} \tag{18}$$

If now we use a suffix 1 on ρ and T to denote general values in the mainstream at the edge of the boundary layer

$$\frac{1}{\rho_1}\frac{\partial p}{\partial x} = -U\frac{dU}{dx}, \tag{19}$$

where U is the mainstream velocity. Hence

$$\frac{1}{\rho}\frac{\partial p}{\partial x} = -\frac{\rho_1}{\rho}U\frac{dU}{dx} = -\frac{T}{T_1}U\frac{dU}{dx},$$

since the pressure is constant across a section of the boundary layer. Finally

$$\frac{1}{p}\frac{\partial p}{\partial x} = -\frac{\rho_1}{p}U\frac{dU}{dx} = -\frac{\gamma}{a_1^2}U\frac{dU}{dx}, \tag{20}$$

where a_1 denotes the local velocity of sound at the edge of the boundary layer. Hence equation (4) takes on the form

$$\frac{\partial^2\chi}{\partial x\partial Y}\frac{\partial\chi}{\partial Y} - \frac{\partial^2\chi}{\partial Y^2}\frac{\partial\chi}{\partial x} = U\frac{dU}{dx}\left[\frac{T}{T_1} - \frac{\gamma}{2a_1^2}\chi\frac{\partial^2\chi}{\partial Y^2}\right] + \nu_s\frac{\partial^3\chi}{\partial Y^3}. \tag{21}$$

The boundary conditions are

$$u = v = 0 \text{ at } y = 0, \quad u \to U \text{ as } y \to \infty,$$

and these imply

$$\chi = \frac{\partial\chi}{\partial Y} = 0 \text{ at } Y = 0, \quad \frac{\partial\chi}{\partial Y} \to U \text{ as } Y \to \infty. \tag{22}$$

Equation (21) together with its boundary conditions may be compared with the equation for the stream function in incompressible flow for the same distribution of mainstream velocity

$$\frac{\partial^2\psi}{\partial x\partial y}\frac{\partial\psi}{\partial y} - \frac{\partial^2\psi}{\partial y^2}\frac{\partial\psi}{\partial x} = U\frac{dU}{\partial x} + \nu\frac{\partial^3\psi}{\partial y^3}, \tag{23}$$

which has boundary conditions

$$\frac{\partial\psi}{\partial x} = \frac{\partial\psi}{\partial y} = 0 \text{ at } y = 0, \quad \frac{\partial\psi}{\partial y} \to U \text{ as } y \to \infty. \tag{24}$$

(The condition $\partial\psi/\partial x = 0$ at $y = 0$ could be replaced by $\psi = 0$ at $y = 0$ without loss.) This flow will be termed the associated incompressible flow.

The effects of compressibility may therefore be thought of as summarized in the term in square brackets

$$G = \frac{T}{T_1} - \frac{\gamma}{2a_1^2}\chi\frac{\partial^2\chi}{\partial Y^2}, \tag{25}$$

which multiplies the effective pressure gradient, and in the altered scale in the direction normal to the boundary.

Except in the case where $dU/dx = 0$ about which some comments will be made later we cannot proceed further without a knowledge of the temperature distribution. This is provided by the energy equation of course.†

---

† Even when $dU/dx = 0$ a knowledge of the temperature distribution is essential before the solution can be completed and y determined from Y.

3. THE ENERGY EQUATION

The energy equation may be written

$$\rho J c_p\left(u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}\right)-u\frac{\partial p}{\partial x} = J\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)+\mu\left(\frac{\partial u}{\partial y}\right)^2, \tag{26}$$

where k is the conductivity. In terms of the Prandtl number σ

$$k = \rho c_p \kappa = \frac{\mu c_p}{\sigma}.$$

If now we multiply equation (4) by ρu and add to (26) we find

$$\rho\left[u\frac{\partial}{\partial x}(\tfrac{1}{2}u^2+Jc_p T)+v\frac{\partial}{\partial y}(\tfrac{1}{2}u^2+Jc_p T)\right]=\frac{\partial}{\partial y}\left[\mu\frac{\partial}{\partial y}\left(\tfrac{1}{2}u^2+\frac{Jc_p T}{\sigma}\right)\right] \tag{27}$$

assuming c_p and σ are constants. Hence, whatever be the form of the temperature variation of μ, if $\sigma = 1$

$$Jc_p T + \tfrac{1}{2}u^2 = \text{const.}$$

is a particular integral of (26) (see Crocco 1946) and is such that

$$\frac{\partial T}{\partial y} = 0 \quad \text{when} \quad u = 0,$$

i.e. it corresponds to a boundary which is thermally insulating. Hence we find

$$Jc_p T+\tfrac{1}{2}u^2 = Jc_p T_1+\tfrac{1}{2}U^2, \tag{28}$$

so that

$$\begin{aligned}\frac{T}{T_1} &= 1+\frac{1}{2Jc_p T_1}(U^2-u^2)\\ &= 1+\frac{R\rho_1}{2Jc_p p}(U^2-u^2)\\ &= 1+\frac{(\gamma-1)}{2a_1^2}(U^2-u^2).\end{aligned} \tag{29}$$

Hence the term G which multiplies $U(dU/dx)$ in equation (21) takes the form

$$G = 1+\frac{(\gamma-1)}{2a_1^2}\left[U^2-\left(\frac{\partial\chi}{\partial Y}\right)^2\right]-\frac{\gamma}{2a_1^2}\chi\frac{\partial^2\chi}{\partial Y^2}, \tag{30}$$

and that the full equation of motion is

$$\frac{\partial^2\chi}{\partial x\partial Y}\frac{\partial\chi}{\partial Y}-\frac{\partial^2\chi}{\partial Y^2}\frac{\partial\chi}{\partial x}=\left[1+\frac{(\gamma-1)}{2a_1^2}\left\{U^2-\left(\frac{\partial\chi}{\partial Y}\right)^2\right\}-\frac{\gamma}{2a_1^2}\chi\frac{\partial^2\chi}{\partial Y^2}\right]U\frac{dU}{dx}+\nu_s\frac{\partial^3\chi}{\partial Y^3}. \tag{31}$$

At the boundary $G = 1+\dfrac{(\gamma-1)}{2a_1^2}U^2$ whilst as $Y\to\infty$, $G\to 1$ as is obviously required. The effect of compressibility is therefore to exaggerate the effect of the pressure gradient in the neighbourhood of the boundary as compared with the associated incompressible flow. In one or two examples which will be discussed later it appears that at least for small Mach numbers G becomes less than unity in the outer part of the boundary layer (i.e. G tends to 1 from below) so that there is a corresponding reduction in effect in the outer parts of the layer.

It also appears that, as Illingworth (unpublished) found, no 'similar' solutions of equation (31) other than the one corresponding to U = const. exist. A number of solutions in series will be considered below but since Pohlhausen's method will also be used we shall now consider the momentum integral equation corresponding to equation (31).

4. Momentum integral equation

Let us denote by δ' the boundary layer thickness measured in terms of Y and write

$$\delta'_1 = \int_0^{\delta'} \left(1 - \frac{u}{U}\right) dY, \tag{32}$$

$$\vartheta' = \int_0^{\delta'} \left(1 - \frac{u}{U}\right) \frac{u}{U}\, dY. \tag{33}$$

Consider first of all the term obtained by integrating G

$$\begin{aligned}\int_0^{\delta'} G\,dY &= \delta' + \frac{(\gamma-1)}{2a_1^2} U^2 \left\{\delta' - \int_0^{\delta'} \frac{u^2}{U^2}\, dY\right\} - \frac{\gamma}{2a_1^2}\left[\chi \frac{\partial\chi}{\partial Y}\right]_0^{\delta'} + \frac{\gamma}{2a_1^2}\int_0^{\delta'} \left(\frac{\partial\chi}{\partial Y}\right)^2 dY \\ &= \delta' + \frac{U^2}{2a_1^2}\int_0^{\delta'} \frac{u^2}{U^2}\, dY - \frac{\gamma}{2a_1^2} U \int_0^{\delta'} \frac{\partial\chi}{\partial Y}\, dY + \frac{(\gamma-1)}{2a_1^2} U^2\delta' \\ &= \delta' + \frac{U^2}{2a_1^2}[\delta' - \delta'_1 - \vartheta'] - \frac{\gamma U^2}{2a_1^2}[\delta' - \delta'_1] + \frac{(\gamma-1)}{2a_1^2} U^2 \delta' \\ &= \delta' - \frac{1}{2}\frac{U^2}{a_1^2}\vartheta' + \frac{(\gamma-1)}{2a_1^2} U^2 \delta'_1.\end{aligned} \tag{34}$$

Hence then exactly as for incompressible flow

$$\frac{\partial}{\partial x}\int_0^{\delta'} u^2\, dY - U \frac{\partial}{\partial x}\int_0^{\delta'} u\, dY = U\frac{dU}{dx}\left[\delta' + \frac{U^2}{2a_1^2}\{(\gamma-1)\,\delta'_1 - \vartheta'\}\right] - \nu_s\left(\frac{\partial u}{\partial Y}\right)_0, \tag{35}$$

i.e. $$-U^2\frac{d\vartheta'}{dx} - U\frac{dU}{dx}(2\vartheta' + \delta'_1 - \delta') = U\frac{dU}{dx}\left[\delta' + \frac{U^2}{2a_1^2}\{(\gamma-1)\,\delta'_1 - \vartheta'\}\right] - \nu_s\left(\frac{\partial u}{\partial Y}\right)_0.$$

Hence $$U^2\frac{d\vartheta'}{dx} + U\frac{dU}{dx}\left[\vartheta'\left(2 - \frac{U^2}{2a_1^2}\right) + \delta'\left\{1 + \frac{(\gamma-1)}{2}\frac{U^2}{a_1^2}\right\}\right] = \nu_s\left(\frac{\partial u}{\partial Y}\right)_0.\dagger \tag{36}$$

This equation may be compared with the associated incompressible form

$$U^2\frac{d\vartheta}{dx} + U\frac{dU}{dx}[2\vartheta + \delta_1] = \nu\left(\frac{\partial u}{\partial y}\right)_0. \tag{37}$$

---

† This equation may be obtained from first principles when due allowance is made for the altered scale in the y-direction, δ'_1 and ϑ' being defined as

$$\delta'_1 = \int_0^{\delta'}\left(1 - \frac{u}{U}\right) dY = \left(\frac{p}{p_s}\right)^{\frac{1}{2}}\int_0^{\delta}\left(1 - \frac{u}{U}\right)\frac{T_s}{T}\, dy = \int_0^{\delta}\left(\frac{\nu_s}{\nu}\right)^{\frac{1}{2}}\left(1 - \frac{u}{U}\right) dy,$$

$$\vartheta' = \int_0^{\delta'}\frac{u}{U}\left(1 - \frac{u}{U}\right) dY = \left(\frac{p}{p_s}\right)^{\frac{1}{2}}\int^{\delta}\frac{u}{U}\left(1 - \frac{u}{U}\right)\frac{T_s}{T}\, dy = \int_0^{\delta}\left(\frac{\nu_s}{\nu}\right)^{\frac{1}{2}}\frac{u}{U}\left(1 - \frac{u}{U}\right) dy.$$

5. THE KÁRMÁN-POHLHAUSEN METHOD

The standard Kármán-Pohlhausen method for incompressible flow is then immediately applicable to equation (36). Let us for the moment write $M_1 = U/a_1$; we notice that if a_0 is the velocity of sound corresponding to rest conditions in the main stream then

$$a_1^2+\frac{(\gamma-1)}{2}U^2 = a_0^2, \tag{38}$$

so that

$$M_1^2 = \frac{U^2}{a_0^2-\frac{1}{2}(\gamma-1)\,U^2} = \frac{M_0^2}{1-\frac{1}{2}(\gamma-1)\,M_0^2}, \tag{39}$$

and

$$1+\frac{(\gamma-1)}{2}M_1^2 = \frac{a_0^2}{a_0^2-\frac{1}{2}(\gamma-1)\,U^2} = \frac{1}{1-\frac{1}{2}(\gamma-1)\,M_0^2}, \tag{40}$$

where $M_0 = U/a_0$.

The theory will be developed with the standard quartic form of velocity distribution putting

$$u/U = f(Y/\delta') = F(\eta)+\lambda G(\eta), \tag{41}$$

where

$$\eta = Y/\delta', \quad F(\eta) = 2\eta-2\eta^3+\eta^4, \quad G(\eta) = \tfrac{1}{6}\eta(1-\eta)^3, \tag{42}$$

and

$$f(0) = 0, \quad f''(0) = -\lambda, \quad f'(1) = f''(1) = 0. \tag{43}$$

Now

$$f''(0) = \frac{\delta'^2}{U}\left(\frac{\partial^2 u}{\partial Y^2}\right)_0 = \frac{\delta'^2}{U}\left(\frac{\partial^3 \chi}{\partial Y^3}\right)_0 = -\frac{\delta'^2}{U\nu_s}\left[1+\frac{(\gamma-1)}{2}M_1^2\right]U\frac{dU}{dx}.$$

Hence

$$\lambda = \frac{\delta'^2}{\nu_s}\frac{dU}{dx}\left[1+\frac{(\gamma-1)}{2}M_1^2\right], \tag{44}$$

and differs from the corresponding quantity for incompressible flow by the factor $[(1+\frac{1}{2}(\gamma-1)\,M_1^2]$. It still remains true as for incompressible flow that

$$\vartheta' = \frac{\delta'}{315}\left(37-\frac{\lambda}{3}-\frac{5\lambda^2}{144}\right), \quad \delta_1' = \frac{\delta'}{120}(36-\lambda) \quad \text{and} \quad \left(\frac{\partial u}{\partial Y}\right)_0 = \frac{U}{\delta'}\left(2+\frac{\lambda}{6}\right). \tag{45}$$

Substituting these values in equation (36) leads to the following differential equation for λ:

$$\frac{d\lambda}{dx} = \frac{d^2U/dx^2}{dU/dx}[\lambda+\lambda^2 h(\lambda)]+\frac{1}{U}\frac{dU}{dx}\left[g(\lambda)+\gamma M_1^2\{\lambda+\lambda^2 h(\lambda)\}+\frac{(\gamma-1)}{2}M_1^2 j(\lambda)\right], \tag{46}$$

where

$$g(\lambda) = \frac{15120-2784\lambda+79\lambda^2+\frac{5}{3}\lambda^3}{(12-\lambda)(37+\frac{25}{12}\lambda)}, \tag{47}$$

$$h(\lambda) = \frac{8+\frac{5}{3}\lambda}{(12-\lambda)(37+\frac{25}{12}\lambda)}, \tag{48}$$

$$j(\lambda) = \frac{15120-1008\lambda+63\lambda^2}{(12-\lambda)(37+\frac{25}{12}\lambda)}. \tag{49}$$

The functions g and h are the same as those occurring in incompressible flow and tabulated in *Modern developments of fluid dynamics*, **1**, 160. The function $j(\lambda)$ is tabulated in table 1 together with

$$k(\lambda) = \frac{(\gamma-1)}{2}j(\lambda)+\gamma\{\lambda^2 h(\lambda)+\lambda\}, \tag{50}$$

when $\gamma = 1{\cdot}4$.

Equation (46) may also be put in a form analogous to the one found most useful in incompressible flow problems by writing

$$\zeta = \lambda \Big/ \frac{dU}{dx}.$$

Then
$$\frac{d\zeta}{dx} = \frac{1}{U}\left[g(\lambda) + \gamma M^2\{\lambda^2 h(\lambda) + \lambda\} + \frac{(\gamma-1)}{2} M^2 j(\lambda)\right] + \zeta^2 \frac{d^2U}{dx^2} h(\lambda). \tag{51}$$

Near a stagnation point the equations reduce to the standard incompressible form and since $M^2/U \to 0$ as $U \to 0$ we find

$$\lambda = 7{\cdot}052, \quad \frac{d\zeta}{dx} = -5{\cdot}391 \frac{d^2U}{dx^2} \Big/ \left(\frac{dU}{dx}\right)^2 \tag{52}$$

for the conditions holding at the stagnation point.

TABLE 1

λ	$j(\lambda)$	$k(\lambda)$	λ	$j(\lambda)$	$k(\lambda)$
12·0	∞	∞	3·0	32·53	11·13
11·0	194·52	128·76	2·0	32·44	9·442
10·0	98·04	63·47	1·0	32·97	8·026
9·0	66·67	41·53	0·0	34·05	6·811
8·0	51·65	30·44	−1·0	35·67	6·753
7·8	49·12	28·74	−2·0	37·83	4·823
7·6	47·74	27·38	−3·0	40·57	3·995
7·4	46·07	26·04	−4·0	43·95	3·255
7·2	44·58	24·81	−5·0	48·09	2·593
7·052	43·57	23·96	−6·0	53·14	2·002
6·8	42·02	22·63	−7·0	59·31	1·480
6·6	40·91	21·02	−8·0	66·92	1·011
6·4	39·91	20·74	−9·0	76·44	0·613
6·2	39·00	19·89	−10·0	88·57	0·297
6·0	38·18	19·09	−11·0	104·44	0·084
5·0	35·11	15·74	−12·0	126·00	0·000
4·0	33·35	13·17			

The condition $\lambda = -12$ is still, of course, the condition which governs separation.

So far we have been concerned with developing the general theory. Its applications to particular examples will now be considered.

6. FLAT PLATE SOLUTION

The solution of the flat plate problem has been given by Crocco (1946) but it seems worthwhile to make a few comments for the sake of completeness.

If $dU/dx = 0$ then equations (21) and (31) reduce to

$$\frac{\partial^2\chi}{\partial x \partial Y}\frac{\partial\chi}{\partial Y} - \frac{\partial^2\chi}{\partial Y^2}\frac{\partial\chi}{\partial Y} = \nu_e \frac{\partial^3\chi}{\partial Y^3}, \tag{53}$$

and the boundary conditions are

$$\frac{\partial\chi}{\partial Y} = 0, \ \chi = 0 \ \text{at} \ Y = 0, \quad \frac{\partial\chi}{\partial Y} \to U \ \text{as} \ Y \to \infty$$

The solution is known (since it is the same as the associated incompressible solution) to be

$$\chi = (\nu_s U x)^{\frac{1}{2}} f(\eta), \quad u = \frac{U}{2} f'(\eta), \tag{54}$$

where $$\eta = \frac{1}{2}\left(\frac{U}{\nu_s x}\right)^{\frac{1}{2}} Y \quad \text{and} \quad f''' + f f'' = 0,$$

with the boundary conditions

$$f = f' = 0 \text{ at } \eta = 0, \quad f' = 2 \text{ at } \eta = \infty.$$

The skin friction τ at the plate is

$$\mu\left(\frac{\partial u}{\partial y}\right)_0 = \frac{\mu_s T}{T_s}\left(\frac{\partial u}{\partial Y}\right)_0 \frac{T_s}{T} = \mu_s\left(\frac{\partial^2 \chi}{\partial Y^2}\right)_0$$

$$= \frac{\mu_s}{4}\frac{U^{\frac{3}{2}}}{(\nu_s x)^{\frac{1}{2}}} f''(0). \tag{55}$$

We can take the standard conditions to refer to the uniform conditions in the main stream and so obtain, using suffix 1 for mainstream values,

$$\tau = \tfrac{1}{4}\alpha \rho_1 U\left(\frac{\nu_1 U}{x}\right)^{\frac{1}{2}}, \tag{56}$$

where $\alpha = 1{\cdot}32824$. The skin friction coefficient

$$\frac{\tau}{\rho_1 U^2} = \tfrac{1}{4}\alpha\left(\frac{\nu_1}{U x}\right)^{\frac{1}{2}} \tag{57}$$

is therefore independent of Mach number in this case.

Although the velocity distribution in terms of Y is also independent of Mach number, this is obviously no longer so in terms of the true scale y. Thus we have u as function of η only where

$$y = \int_0^Y \frac{T}{T_1}\, dY$$

$$= \int_0^Y \left[1 + \frac{(\gamma - 1)}{2a_1^2}(U^2 - u^2)\right] dY$$

$$= \left[1 + \frac{(\gamma-1)}{2} M_1^2\right] Y - \frac{(\gamma - 1)}{2} M_1^2 \int_0^Y \frac{u^2}{U^2}\, dY, \tag{58}$$

so that $$y = \frac{1}{2}\left(\frac{U}{\nu_1 x}\right)^{-\frac{1}{2}}\left[\eta + \frac{(\gamma-1)}{2} M_1^2 \left\{\eta - \frac{1}{4}\int_0^\eta f'^2\, d\eta\right\}\right].$$

Now if we integrate the equation for f we find

$$\int_0^\eta f'^2\, d\eta = f'' - f''(0) + f f'.$$

Hence $$y = \frac{1}{2}\left(\frac{U}{\nu_1 x}\right)^{-\frac{1}{2}}\left[\eta + \frac{(\gamma - 1)}{2} M_1^2 \{\eta - \tfrac{1}{4} f'' - \tfrac{1}{4} f f' + 0{\cdot}34206\}\right], \tag{59}$$

so that if we use y_i to denote the ordinate of the incompressible flow with viscosity ν_1,

$$y_i = \frac{1}{2}\left(\frac{U}{\nu_1 x}\right)^{-\frac{1}{2}}\eta \quad \text{and} \quad \frac{y-y_i}{y_i} = \frac{(\gamma-1)}{2}M_1^2\left\{1-\frac{(f''+ff')}{4\eta}+\frac{0{\cdot}324206}{\eta}\right\}. \tag{60}$$

When $\eta \to 0$ the expression in curly brackets tends to 1. When η is large we find

$$\frac{y-y_i}{y_i} = 0{\cdot}5962\frac{(\gamma-1)}{\eta}M_1^2. \tag{61}$$

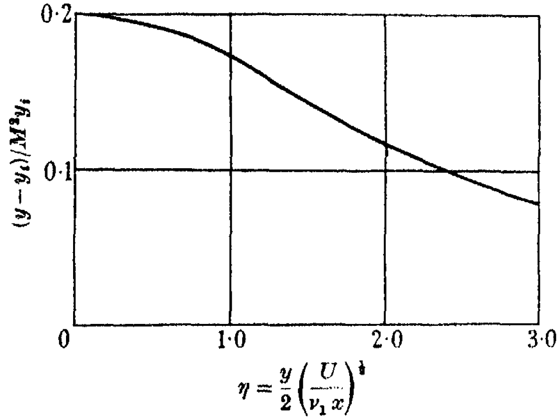


FIGURE 1

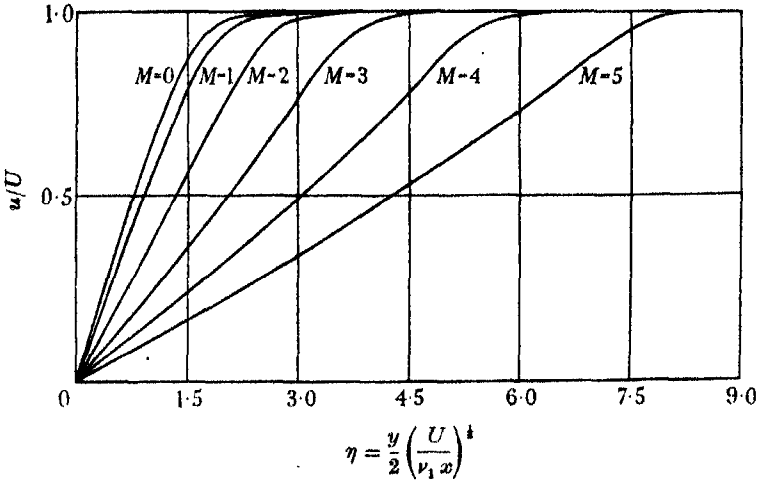


FIGURE 2

To see the significance of this last result it is convenient to imagine a finite boundary layer thickness δ defined by the condition $u/U = 0{\cdot}999$, say. This corresponds to a value of 3 for η. Hence

$$\begin{aligned}\frac{\delta-\delta_i}{\delta_i} &= 0{\cdot}1987(\gamma-1)M_1^2\\ &= 0{\cdot}0795M_1^2\end{aligned} \tag{62}$$

when $\gamma = 1{\cdot}4$. Thus there is an 8 % increase in boundary layer thickness when $M_1 = 1$ and a doubling in thickness when $M_1 = 3{\cdot}5$.

The relation defined by equation (60) is shown graphically in figure 1 when $\gamma = 1{\cdot}4$. Figure 2 contains a number of velocity distributions at different Mach numbers.

The forms of the velocity distributions are much the same as those obtained by Kármán & Tsien (1938) who assumed $\mu \propto T^{0{\cdot}76}$. They found as did Brainerd & Emmons (1941, 1942) with a variety of viscosity-temperature variations that the skin friction decreased slowly as the Mach number increased. The constant value found in the present investigation is due, as suggested earlier, to our assumptions effectively overestimating the effects of viscosity.

7. FLOW IN WAKES AND JETS

Although the problem of viscous compressible flow in wakes and jets may not be very significant practically since the flow is probably turbulent it is nevertheless of some interest to see the effects of compressibility even in this case.

With the usual approximations equation (31) adapted to flow in a wake becomes

$$U\frac{\partial^2\chi}{\partial x\partial Y} = \nu_1\frac{\partial^3\chi}{\partial Y^3}, \tag{63}$$

where ν_1 is the kinematic viscosity corresponding to the undisturbed flow, the drag D being given by the integral

$$D = \rho_1 U^2\int_{-\infty}^{\infty}\frac{\rho}{\rho_1}\frac{u}{U}\left(1-\frac{u}{U}\right)dy. \tag{64}$$

Thus

$$D = \rho_1 U^2\int_{-\infty}^{\infty}\frac{\rho T}{\rho_1 T_1}\frac{u}{U}\left(1-\frac{u}{U}\right)dY$$

$$= \rho_1 U^2\int_{-\infty}^{\infty}\frac{u}{U}\left(1-\frac{u}{U}\right)dY. \tag{65}$$

Putting $w = 1-u/U$ we then have

$$D = \rho_1 U^2\int_{-\infty}^{\infty} w\,dY \tag{66}$$

approximately.

The solution in terms of Y is then given by the standard incompressible solution

$$w = w_0 e^{-\frac{1}{2}\eta^2}, \tag{67}$$

where

$$\eta = \left(\frac{U}{2\nu_1 x}\right)^{\frac{1}{2}} Y \quad\text{and}\quad w_0 = \frac{D}{2\rho_1 U^2}\left(\frac{U}{\pi\nu_1 x}\right)^{\frac{1}{2}}. \tag{68}$$

The integral

$$Jc_p T + \tfrac{1}{2}u^2 = \text{const.}$$

of the energy equation is applicable to this problem too since it makes $\partial T/\partial y$ vanish with $\partial u/\partial y$. Hence as in the flat plate problem we have a velocity distribution determined by η, the true scale y being given by

$$y = Y + \frac{(\gamma-1)}{2} M_1^2 \int_0^Y \left(1 - \frac{u^2}{U^2}\right) dY$$

$$= Y + (\gamma-1) M_1^2 \int_0^Y w\,dY. \tag{69}$$

Hence if Y is sufficiently large

$$y = Y + \frac{(\gamma-1) M_1^2 D}{2\rho_1 U^2}. \tag{70}$$

As for the flow along a flat plate we can define a width of wake by $w/w_0 = 0{\cdot}1\ \%$, say, which gives a value of Y of $\frac{4(2\nu_1 x)^{\frac{1}{2}}}{U}$ approximately.

Of course $D/\rho_1 U^2$ is in general an unknown function of both Reynolds and Mach numbers for a given obstacle so that variation in wake size implied by (70) is also not known. One can, however, interpret (70) by saying that for a given value of $D/\rho_1 U^2$ the ordinary incompressible theory with kinematic viscosity ν_1 would underestimate the wake thickness by a constant amount $\frac{(\gamma-1)}{2} M_1^2 \frac{D}{\rho_1 U^2}$.

In the particular case of the wake behind a flat plate of length l

$$\frac{D}{\rho_1 U^2} = 1{\cdot}328 l \left(\frac{Ul}{\nu_1}\right)^{-\frac{1}{2}}, \tag{71}$$

and

$$\frac{u}{U} = 1 - \frac{0{\cdot}664}{\sqrt{\pi}} \left(\frac{l}{x}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\eta^2}, \tag{72}$$

where

$$\eta = \left(\frac{U}{2\nu_1 x}\right)^{\frac{1}{2}} Y, \tag{73}$$

and

$$y = Y + 0{\cdot}664(\gamma-1) M_1^2 l \left(\frac{Ul}{\nu_1}\right)^{-\frac{1}{2}} \operatorname{erf}\left[\frac{Y}{2}\left(\frac{U}{\nu_1 x}\right)^{\frac{1}{2}}\right], \tag{74}$$

where

$$\operatorname{erf} t = \frac{2}{\sqrt{\pi}} \int_0^t e^{-t^2} dt.$$

Defining the edge of the wake as in the general case above the implied total width of the wake is

$$8\left(\frac{2\nu_1 x}{U}\right)^{\frac{1}{2}} + 1{\cdot}328(\gamma-1) M_1^2 l \left(\frac{\nu_1}{Ul}\right)^{\frac{1}{2}} = l\left(\frac{\nu_1}{Ul}\right)^{\frac{1}{2}} \left[8\sqrt{2}\left(\frac{x}{l}\right)^{\frac{1}{2}} + 1{\cdot}328(\gamma-1) M_1^2\right]$$

$$= l\left(\frac{\nu_1}{Ul}\right)^{\frac{1}{2}} \left[11{\cdot}314\left(\frac{x}{l}\right)^{\frac{1}{2}} + 0{\cdot}5313 M_1^2\right], \tag{75}$$

when $\gamma = 1{\cdot}4$.

The velocity distribution in the wake when $x = 10l$ for Mach numbers of 0, 5 and 10 is shown graphically in figure 3.

Consider next a two-dimensional jet. The flux of momentum F across each section of the jet is

$$F = 2\int_0^\infty \rho u^2 dy$$
$$= 2\rho_0 \int_0^\infty \frac{\rho}{\rho_0} u^2 dy = 2\rho_0 \int_0^\infty u^2 dY, \tag{76}$$

where the suffix 0 refers to conditions in the gas at rest outside the jet.

The equation of motion derived from (31) is

$$\frac{\partial^2\chi}{\partial x \partial Y}\frac{\partial\chi}{\partial Y} - \frac{\partial^2\chi}{\partial Y^2}\frac{\partial\chi}{\partial x} = \nu_0 \frac{\partial^3\chi}{\partial Y^3}. \tag{77}$$

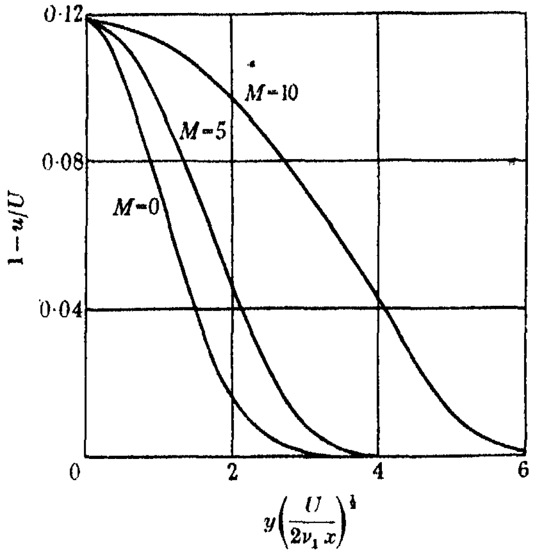


Figure 3

The standard incompressible solution can then be used to give

$$u = \frac{\partial\chi}{\partial Y} = 0{\cdot}4543\left(\frac{F^2}{\rho_0^2 \nu_0 x}\right)^{\frac{1}{3}} \operatorname{sech}^2 \xi, \tag{78}$$

where
$$\xi = 0{\cdot}2752\left(\frac{F}{\rho_0 \nu_0^2 x^2}\right)^{\frac{1}{3}} Y. \tag{79}$$

The ordinate y is then given by

$$y = Y - \frac{(\gamma-1)}{2a_0^2}\int_0^Y u^2 dY, \tag{80}$$

where a_0 is the velocity of sound corresponding to rest conditions. For sufficiently large Y we see that $y = Y - \frac{(\gamma-1)}{4a_0^2}\frac{F}{\rho_0}$, so that for a given value of F/ρ_0 the jet width is less by an amount $\frac{(\gamma-1)}{2a_0^2}\frac{F}{\rho_0}$ than that obtained by incompressible theory with kinematic viscosity ν_0.

Of course the considerations set out in the section on wakes and jets do not hold in the neighbourhood of $x = 0$ where neither the assumptions of the boundary layer nor the neglect of pressure gradient are valid.

8. FLOW WITH LINEARLY INCREASING VELOCITY IN THE MAIN STREAM

One of the standard solutions of incompressible flow is given by considering the velocity distribution

$$U = \beta x$$

in the mainstream. It corresponds to flow in the vicinity of a stagnation point. It is obvious of course that even when the flow is compressible the velocity distribution in the boundary layer in the immediate vicinity of the stagnation point must be the same as the incompressible one. If, however, the linear velocity distribution $U = \beta x$ may be supposed to hold for some appreciable distance downstream then compressibility will begin to play a part. To this extent the problem is more artificial than the corresponding incompressible problem but it is worthy of consideration since it may be typical of accelerated flow problems generally.

With $U = \beta x$ equation (31) takes the form

$$\frac{\partial^2\chi}{\partial x\,\partial Y}\frac{\partial\chi}{\partial Y} - \frac{\partial^2\chi}{\partial Y^2}\frac{\partial\chi}{\partial x} = \left[1 + \frac{(\gamma-1)}{2a_1^2}\left\{\beta^2x^2 - \left(\frac{\partial\chi}{\partial Y}\right)^2\right\} - \frac{\gamma}{2a_1^2}\chi\frac{\partial^2\chi}{\partial Y^2}\right]\beta^2 x + \nu_0\frac{\partial^3\chi}{\partial Y^3}, \tag{81}$$

where we use the suffix zero to denote conditions at the stagnation point and treat these as our standard.† We see that

$$a_1^2 + \frac{(\gamma-1)}{2}\beta^2x^2 = a_0^2, \tag{82}$$

so that

$$\frac{1}{a_1^2} = \frac{1}{a_0^2}\left[1 + \frac{(\gamma-1)}{2}\frac{\beta^2x^2}{a_0^2} + \left(\frac{\gamma-1}{2}\right)^2\frac{\beta^4x^4}{a_0^4} + \dots\right]. \tag{83}$$

We can obtain a series solution in the form

$$\chi = (\nu_0\beta)^{\frac{1}{2}}x\left[f_1(\eta) + \frac{\beta^2x^2}{a_0^2}f_3(\eta) + \frac{\beta^4x^4}{a_0^4}f_5(\eta) + \dots\right], \tag{84}$$

where

$$\eta = \left(\frac{\beta}{\nu_0}\right)^{\frac{1}{2}}Y. \tag{85}$$

The differential equations satisfied by f_1, f_3 and f_5 are

$$f_1'^2 - f_1f_1'' = 1 + f_1''', \tag{86}$$

$$4f_1'f_3' - 3f_1''f_3 - f_1f_3'' = \frac{(\gamma-1)}{2}(1 - f_1'^2) - \frac{\gamma}{2}f_1f_1'' + f_3''', \tag{87}$$

$$\begin{aligned}6f_1'f_5' - 5f_1''f_5 - f_1f_5'' = {} & \frac{(\gamma-1)^2}{4}(1 - f_1'^2) - \gamma\frac{(\gamma-1)}{4}f_1f_1'' \\ & - (\gamma-1)f_1'f_3' - \frac{\gamma}{2}(f_1f_3'' + f_3f_1'') - 3(f_3'^2 - f_3f_3'') + f_5''',\end{aligned} \tag{88}$$

---

† The suffix 1 on a is used as before to denote general mainstream values.

and the boundary conditions are

$$f_1 = f_1' = f_3 = f_3' = f_5 = f_5' = 0 \text{ at } \eta = 0, \quad f_1' = 1, f_3' = f_5' = 0 \text{ at } \eta = \infty.$$

As remarked above f_1 is the same function that appears in the solution of the incompressible problem. Taking $\gamma = 1{\cdot}4$ we have

$$4f_1'f_3' - 3f_1''f_3 - f_1f_3'' = 0{\cdot}2(1 - f_1'^2) - 0{\cdot}7f_1f_1'' + f_3''', \tag{89}$$

$$\begin{aligned} 6f_1'f_5' - 5f_1''f_5 - f_1f_5'' = {} & 0{\cdot}04(1 - f_1'^2) - 0{\cdot}14f_1f_1'' - 0{\cdot}4f_1'f_3' \\ & - 0{\cdot}7(f_1f_3'' + f_3f_1'') - 3(f_3'^2 - f_3f_3'') + f_5'''. \end{aligned} \tag{90}$$

The velocity distribution is given by

$$u = \frac{\partial \chi}{\partial Y} = \beta x\left[f_1'(\eta) + \frac{\beta^2x^2}{a_0^2}f_3'(\eta) + \frac{\beta^4x^4}{a_0^4}f_5'(\eta) + \dots\right], \tag{91}$$

where $$\eta = \left(\frac{\beta}{\nu_0}\right)^{\frac{1}{2}} Y$$

and $$\begin{aligned} y &= \left(\frac{p_0}{p}\right)^{\frac{1}{2}} \int_0^Y \frac{T}{T_0}\, dY \\ &= \left(\frac{p_0}{p}\right)^{\frac{1}{2}} \left[Y - \frac{(\gamma-1)}{2a_0^2}\int_0^Y u^2\, dY\right] \\ &= \left(\frac{p_0}{p}\right)^{\frac{1}{2}} \left(\frac{\nu_0}{\beta}\right)^{\frac{1}{2}} \left[\eta - \frac{(\gamma-1)}{2a_0^2}\beta^2x^2\int_0^\eta f_1'^2\, d\eta - (\gamma-1)\frac{\beta^4x^4}{a_0^4}\int_0^\eta f_1'f_3'\, d\eta + \dots\right]. \end{aligned} \tag{92}$$

The skin friction τ at the wall is given by

$$\tau = \mu_w\left(\frac{\partial u}{\partial y}\right)_w = \frac{\mu_w T_0}{T_w}\left(\frac{p}{p_0}\right)^{\frac{1}{2}}\left(\frac{\partial u}{\partial Y}\right)_w, \tag{93}$$

where the suffix w refers to conditions at the wall. Hence

$$\tau = \rho_0\nu_0^{\frac{1}{2}}\left(\frac{p}{p_0}\right)^{\frac{1}{2}}\beta^{\frac{3}{2}}x\left[f_1''(0) + \frac{\beta^2x^2}{a_0^2}f_3''(0) + \frac{\beta^4x^4}{a_0^4}f_5''(0) + \dots\right]. \tag{94}$$

The pressure ratio p/p_0 which appears in equations (92) and (94) is given by

$$\frac{p}{p_0} = \left[1 - \frac{(\gamma-1)}{2}\frac{\beta^2x^2}{a_0^2}\right]^{\gamma/(\gamma-1)}, \tag{95}$$

and may be expanded in powers of β^2x^2/a_0^2 to give

$$\begin{aligned} y = \left(\frac{\nu_0}{\beta}\right)^{\frac{1}{2}} & \left[\eta + \frac{\beta^2x^2}{a_0^2}\left\{\frac{\gamma\eta}{4} - \frac{(\gamma-1)}{2}\int_0^\eta f_1'^2\, d\eta\right\}\right. \\ & \left. + \frac{\beta^4x^4}{a_0^4}\left\{\gamma\frac{(3\gamma-2)}{32}\eta - \gamma\frac{(\gamma-1)}{8}\int_0^\eta f_1'^2\, d\eta - (\gamma-1)\int_0^\eta f_1'f_3'\, d\eta\right\} + \dots\right], \end{aligned} \tag{96}$$

and $$\begin{aligned} \tau = \rho_0\nu_0^{\frac{1}{2}}\beta^{\frac{3}{2}}x & \left[f_1''(0) + \frac{\beta^2x^2}{a_0^2}\left\{f_3''(0) - \frac{\gamma}{4}f_1''(0)\right\}\right. \\ & \left. + \frac{\beta^4x^4}{a_0^4}\left\{f_5''(0) - \frac{\gamma}{4}f_3''(0) + \gamma\frac{(2-\gamma)}{32}f_1''(0)\right\} + \dots\right]. \end{aligned} \tag{97}$$

The integration of equation (89) for f_3 has been carried out and the results are shown in table 2; f_3' is shown graphically in figure 4. This function is in fact quite small compared with f_1; $f_3''(0)$ is positive and of the order of 5 % of $f_1''(0)$. The velocity correction implied by f_3' is positive in the inner quarter, say, of the boundary layer with a maximum of just less than 1 % of the maximum value, unity, of f_1' and is negative in the remainder with a maximum numerical value of just over 1 % of the maximum value of f_1'.

TABLE 2

η	f_3	f_3'	f_3''	η	f_3	f_3'	f_3''
0·0	0·000	0·000	0·056	1·6	−0·003	−0·012	0·002
0·1	0·000	0·005	0·036	1·7	−0·004	−0·011	0·005
0·2	0·001	0·007	0·018	1·8	−0·005	−0·011	0·007
0·3	0·002	0·008	0·003	1·9	−0·006	−0·010	0·009
0·4	0·003	0·008	−0·010	2·0	−0·007	−0·009	0·010
0·5	0·003	0·007	−0·018	2·1	−0·008	−0·008	0·010
0·6	0·004	0·004	−0·024	2·2	−0·008	−0·007	0·010
0·7	0·004	0·002	−0·027	2·3	−0·009	−0·006	0·010
0·8	0·004	−0·001	−0·027	2·4	−0·009	−0·005	0·009
0·9	0·004	−0·004	−0·025	2·5	−0·010	−0·004	0·008
1·0	0·003	−0·006	−0·022	2·6	−0·010	−0·003	0·007
1·1	0·003	−0·008	−0·018	2·7	−0·011	−0·003	0·007
1·2	0·002	−0·010	−0·014	2·8	−0·011	−0·002	0·006
1·3	0·001	−0·011	−0·009	2·9	−0·011	−0·002	0·005
1·4	0·000	−0·011	−0·005	3·0	−0·011	−0·001	0·004
1·5	−0·001	−0·012	−0·001				

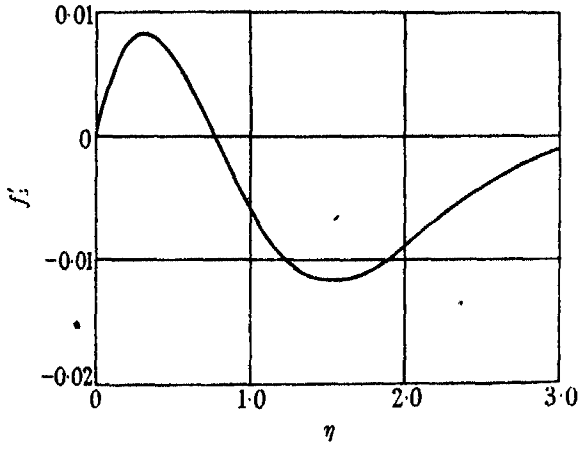


FIGURE 4

The integration of equation (90) has not been carried out but a crude examination indicates that f_5 should have much the same form as f_3 but considerably smaller in magnitude possibly of the order of 20 % of f_3.

It appears therefore that even when $\beta x/a_0$ has reached unity the difference in u at a given value of Y (or η) from its corresponding incompressible value is small and of the order of 1 %. This does not of course imply that the change in u at a given value of the true ordinate y is small. In fact the change in scale is much the most important effect here. Bearing in mind the remarks on order of magnitude made above we can write equation (92) as

$$y = \left(\frac{p_0}{p}\right)^{\frac{1}{2}} \left(\frac{\nu_0}{\beta}\right)^{\frac{1}{2}} \left[\eta - \frac{(\gamma-1)}{2a_0^2}\beta^2 x^2 \int_0^\eta f_1'^2\, d\eta\right]$$

approximately. Thus

$$y = \left(\frac{p_0}{p}\right)^{\frac{1}{2}} \left(\frac{\nu_0}{\beta}\right)^{\frac{1}{2}} \left[\eta - \frac{(\gamma-1)}{4a_0^2}\beta^2 x^2 \{\eta + f_1'' + f_1 f_1' - f_1''(0)\}\right]. \tag{98}$$

Velocity distributions corresponding to $\beta x/a_0 = 0$, 0·5 and 1·0 (beyond which stage it seems inadvisable to proceed without calculating f_5) are shown in figure 5.

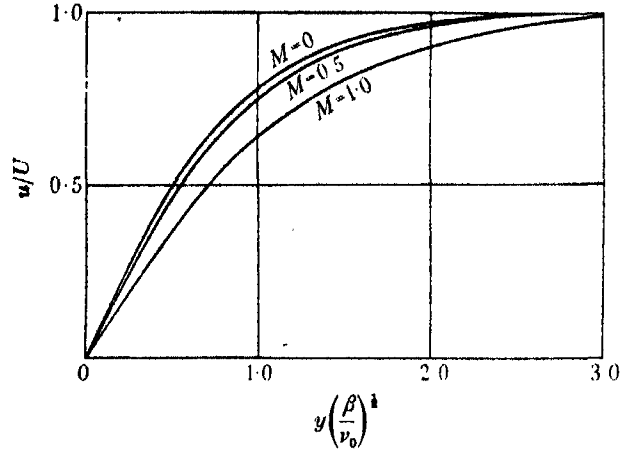


FIGURE 5

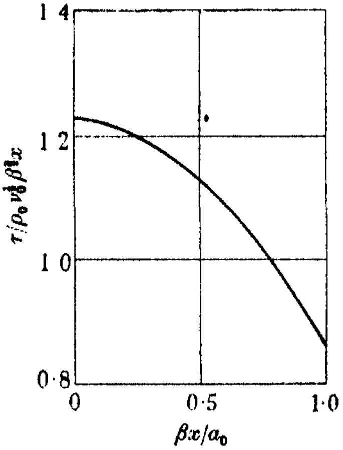


FIGURE 6

The skin friction at the wall τ given by equation (94) is shown graphically in figure 6 for values of $\beta x/a_0$ between 0 and 1. Although $\nu_0^{\frac{1}{2}}\beta^{-\frac{3}{2}}x^{-1}(\partial u/\partial Y)_0$ increases with Mach number from 1·23 at $\beta x/a_0 = 0$ to 1·29 at $\beta x/a_0 = 1$ the actual skin friction coefficient $\tau/\rho_0 \nu_0^{\frac{1}{2}}\beta^{\frac{3}{2}}x$ falls from 1·23 to 0·87 as a consequence of the changing scale. As one would indeed expect in the light of the velocity distributions the factor $(p/p_0)^{\frac{1}{2}}$ representing the effect of change of scale predominates and leads to this substantial decrease in skin friction coefficient with Mach number. A good approximation to the skin friction is, in fact, provided by reducing the incompressible value by the factor $(p/p_0)^{\frac{1}{2}}$ in this range of variation of $\beta x/a_0$.

A problem rather similar to the one discussed here is provided by the boundary layer along the wall of a converging channel. Here the effects of compressibility on the main flow are known and to this extent the problem is less artificial than that of the present paragraph.

9. SERIES EXPANSIONS FOR THE GENERAL MAINSTREAM VELOCITY DISTRIBUTION

The considerations of the previous section make it evident that if the velocity distribution in the main stream is given as a series of ascending powers of x

$$U = \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots \tag{99}$$

a solution could be obtained in the form

$$\chi = (\nu_0\beta_1)^{\frac{1}{2}}\,x[F_1(\eta)+xF_2(\eta)+x^2F_3(\eta)+\dots], \tag{100}$$

where

$$\eta = \left(\frac{\beta_1}{\nu_0}\right)^{\frac{1}{2}} Y. \tag{101}$$

By writing

$$\left.\begin{aligned} F_1 &= f_1, \quad F_2 = \frac{3\beta_2}{\beta_1}f_2, \\ F_3 &= \frac{4\beta_3}{\beta_1}g_3+\frac{4\beta_2^2}{\beta_1^2}h_3+\frac{\beta_1^2}{a_0^2}k_3, \\ F_4 &= \frac{5\beta_4}{\beta_1}g_4+\frac{5\beta_2\beta_3}{\beta_1^2}h_4+\frac{5\beta_2^3}{\beta_1^3}k_4+\frac{\beta_1\beta_2}{a_0^2}q_4, \quad \text{etc.,}\end{aligned}\right\} \tag{102}$$

the equations for the functions F can be split up into equations independent of both Reynolds and Mach numbers, just as in the corresponding problem for incompressible flow.

The functions f_1, f_2, g_3, h_3, g_4, h_4, k_4 are in fact the functions derived for incompressible flow (Howarth 1934). The new functions k_3, q_4 introduced by compressibility are given by the equations

$$4f_1'k_3'-3f_1''k_3-f_1k_3'' = \frac{(\gamma-1)}{2}(1-f_1'^2)-\frac{\gamma}{2}f_1f_1''+k_3''' \tag{103}$$

and

$$\begin{aligned} 5f_1'q_4'-4f_1''q_4-f_1q_4'' &= q_4'''-3(5f_2'k_3'-3f_2''k_3-2f_2k_3'') \\ &\quad +(\gamma-1)[(1-3f_1'f_2')+3(1-f_1'^2)]-\frac{3\gamma}{2}(f_1f_2''+f_1''f_2+f_1f_1''), \end{aligned} \tag{104}$$

for which the boundary conditions are

$$k_3 = k_3' = q_4 = q_4' = 0 \text{ at } \eta = 0, \quad k_3' = q_4' = 0 \text{ at } \eta = \infty.$$

As might be expected equation (103) is identical with (87) of which the solution is given in table 2 (where it is called f_3) when $\gamma = 1{\cdot}4$. The functions f_1, f_2, g_3, h_3 and k_4 are tabulated by Howarth (1934).

In view of the generally slow convergence of the series involved and the consequently limited applicability of the results it has not been considered worthwhile to calculate the compressibility effects beyond the function k_3 already calculated.

The skin friction at the wall is

$$\begin{aligned} \mu_w\left(\frac{\partial u}{\partial y}\right)_w &= \mu_0\left(\frac{\partial u}{\partial Y}\right)_w\left(\frac{p}{p_0}\right)^{\frac{1}{2}} \\ &= \rho_0\nu_0^{\frac{1}{2}}\beta_1^{\frac{3}{2}}\left(\frac{p}{p_0}\right)^{\frac{1}{2}} x\left[f_1''(0)+\frac{3\beta_2x}{\beta_1}f_2''(0)+\left\{\frac{4\beta_3}{\beta_1}g_3''(0)+\frac{4\beta_2^2}{\beta_1^2}h_3''(0)\frac{\beta_1^2}{2a_0^2}k_3''(0)\right\}x^2\right. \\ &\quad \left.+\left\{\frac{5\beta_4}{\beta_1}g_4''(0)+\frac{5\beta_2\beta_3}{\beta_1^2}h_4''(0)+\frac{5\beta_2^3}{\beta_1^3}k_4''(0)+\frac{\beta_1\beta_2}{2a_0^2}q_4''(0)\right\}x^3+\dots\right]. \end{aligned} \tag{105}$$

A similar method applies to velocity distribution

$$U = \beta_1x+\beta_3x^3+\beta_5x^5+\dots \tag{106}$$

corresponding to a symmetrical obstacle.

The stream function χ is given by a series of the form

$$\chi = (\nu_0\beta_1)^{\frac{1}{2}} x[F_1(\eta) + x^2 F_3(\eta) + x^4 F_5(\eta) + \dots], \tag{107}$$

where

$$\eta = \left(\frac{\beta_1}{\nu_0}\right)^{\frac{1}{2}} Y. \tag{108}$$

$$\left.\begin{aligned} F_1 &= f_1, \\ F_3 &= \frac{4\beta_3}{\beta_1} f_3 + \frac{\beta_1^2}{2a_0^2} g_3, \\ F_5 &= \frac{6\beta_5}{\beta_1} g_5 + \frac{6\beta_3^2}{\beta_1^2} h_5 + \frac{\beta_1\beta_3}{2a_0^2} k_5 + \frac{\beta_1^4}{2a_0^4} q^5, \end{aligned}\right\} \tag{109}$$

Here again the functions f_1, f_3, g_5, h_5 are the ones occurring in incompressible flow (Howarth 1934, where they are tabulated). The equations determining g_3, k_5 and q_5 are

$$4f_1' g_3' - 3f_1'' g_3 - f_1 g_3'' = g_3''' + \frac{(\gamma - 1)}{2}(1 - f_1'^2) - \frac{\gamma}{2} f_1 f_1'', \tag{110}$$

$$\begin{aligned} 6f_1' k_5' - 5f_1'' k_5 - f_1 k_5'' = k_5''' &- 12[2f_3' g_3' - f_1'' g_3 - f_1 g_3''] \\ &+ (\gamma - 1)\{1 - 4f_1' f_3'\} - 2\gamma(f_1 f_3'' + f_1'' f_3), \end{aligned} \tag{111}$$

$$\begin{aligned} 6f_1' q_5' - 5f_1'' q_5 - f_1 q_5'' = q_5''' &- 3(g_3'^2 - g_3 g_3'') - (\gamma - 1) f_1' g_3' \\ &- \frac{\gamma}{2}(f_1' g_3'' + f_1 g_3'') + \gamma\frac{(\gamma - 1)}{4}(1 - f_1'^2) - \gamma\frac{(\gamma - 1)}{4} f_1 f_1'', \end{aligned} \tag{112}$$

with the boundary conditions

$$g_3 = g_3' = k_5 = k_5' = q_5 = q_5' = 0 \text{ at } \eta = 0, \quad g_3' = k_5' = q_5' = 0 \text{ at } \eta = \infty.$$

Equation (110) is identical with (103) and (87) the solution being given in table 2 (where it is called f_3). Equation (112) is identical with (88).

Beyond this stage the number of new functions introduced by compressibility effects begins to increase rapidly and the value of the method except at small Mach numbers is doubtful. For this reason the integration has not been carried beyond g_3.

10. THE FLAT PLATE WITH A RETARDING LINEAR VELOCITY GRADIENT

A problem which has been thoroughly investigated in incompressible flow is that which occurs when

$$U = U_0 - U_1 x. \tag{113}$$

This velocity distribution has been shown to lead to separation of the layer when

$$\frac{U_1 x}{U_0} = 0{\cdot}120.$$

We can apply series methods similar to the ones developed above to determine the effects of compressibility. We take conditions at the leading edge as our standard and denote them by a suffix 0. Then we can find a solution of equation (31) in the form

$$\chi = U_0^{\frac{1}{2}} x^{\frac{1}{2}} \nu_0^{\frac{1}{2}}\{F_0(\eta) - 8x^* F_1(\eta) + (8x^*)^2 F_2(\eta) - \dots\}, \tag{114}$$

where
$$\eta = \frac{1}{2}\frac{YU_0^{\frac{1}{2}}}{\nu_0^{\frac{1}{2}}x^{\frac{1}{2}}} \quad \text{and} \quad x^* = \frac{U_1 x}{U_0}. \tag{115}$$

If we put
$$\left.\begin{aligned} F_0 &= f_0, \\ F_1 &= f_1 + M_0^2 g_1, \\ F_2 &= f_2 + M_0^2 g_2 + M_0^4 h_2, \text{ etc.,} \end{aligned}\right\} \tag{116}$$

where M_0 is the Mach number at the leading edge we find that $f_0, f_1, f_2, \ldots$, are the functions occurring in the incompressible solution and given by Howarth (1938). The equations to determine g_1, g_2, h_2 are

$$g_1''' + f_0 g_1'' - 2f_0' g_1' + 3f_0'' g_1 = -\frac{(\gamma-1)}{8}(4 - f_0'^2) + \frac{\gamma}{8} f_0 f_0'', \tag{117}$$

$$\begin{aligned} g_2''' + f_0 g_2'' - 4f_0' g_2' + 5f_0'' g_2 = {} & -\frac{(\gamma-1)}{64}(4 - f_0'^2) + \frac{\gamma}{64} f_0 f_0'' - \frac{(\gamma-1)}{8}(1 - 2f_0' f_1') \\ & + \frac{\gamma}{8}(f_0'' f_1 + f_1'' f_0) + 4f_1' g_1' - 3f_1 g_1'' - 3f_1'' g_1, \end{aligned} \tag{118}$$

$$\begin{aligned} h_2''' + f_0 h_2'' - 4f_0' h_2' + 5f_0'' h_2 = {} & -\frac{(\gamma-1)^2}{64}(4 - f_0'^2) + \gamma\frac{(\gamma-1)}{64} f_0 f_0'' \\ & + \frac{(\gamma-1)}{4} f_0' g_1' + \frac{\gamma}{8}(f_0'' g_1 + f_0 g_1'') + 2g_1'^2 - 3g_1 g_1''. \end{aligned} \tag{119}$$

The boundary conditions are

$$g_1 = g_1' = g_2 = g_2' = h_2 = h_2' = 0 \text{ at } \eta = 0, \quad g_1' = g_2' = h_2' = 0 \text{ at } \eta = \infty.$$

The velocity distribution is given by

$$u = \frac{U_0}{2}[f_0'(\eta) - 8x^*\{f_1'(\eta) + M_0^2 g_1'(\eta)\} + 8x^{*2}\{f_2'(\eta) + M_0^2 g_2(\eta) + M_0^4 h_2(\eta)\} + \ldots]. \tag{120}$$

The ordinate y is determined by

$$\begin{aligned} y &= \left(\frac{p_0}{p}\right)^{\frac{1}{\gamma}} \int_0^Y \frac{T}{T_0}\,dY \\ &= \left(\frac{p_0}{p}\right)^{\frac{1}{\gamma}} \left[Y + \frac{(\gamma-1)}{2} M_0^2 \int_0^Y \left(1 - \frac{u^2}{U_0^2}\right) dY\right. \\ &= \left(\frac{p_0}{p}\right)^{\frac{1}{\gamma}} \frac{2\nu_0^{\frac{1}{2}} x^{\frac{1}{2}}}{U_0^{\frac{1}{2}}} \left[\eta + \frac{(\gamma-1)}{2} M_0^2 \int_0^\eta \left(1 - \frac{u^2}{U_0^2}\right) d\eta\right], \end{aligned} \tag{121}$$

and
$$\frac{p}{p_0} = \left[1 + \frac{(\gamma-1)}{8} M_0^2 \{8x^* - \tfrac{1}{16}(8x^*)^2\}\right]^{\gamma/(\gamma-1)}. \tag{122}$$

The skin friction is given by

$$\begin{aligned} \mu_w\left(\frac{\partial u}{\partial Y}\right)_w = \mu_0\left(\frac{p}{p_0}\right)^{\frac{1}{\gamma}}\left(\frac{\partial u}{\partial Y}\right)_w = \frac{\rho_0 \nu_0^{\frac{1}{2}} U_0^{\frac{3}{2}}}{4}\left(\frac{p}{p_0}\right)^{\frac{1}{\gamma}} & [f_0''(0) - (8x^*)\{f_1''(0) + M_0^2 g_1''(0)\} \\ & + (8x^*)^2\{f_2''(0) + M_0^2 g_2''(0) + M_0^4 h_2''(0)\} + \ldots]. \end{aligned} \tag{123}$$

The new functions introduced by compressibility effects again multiply rapidly. It is to be noted that each of the functions F after F_0 contains a term in M_0^2 so that the compressibility correction at small Mach numbers is correspondingly more difficult to assess than in the accelerated flow problem discussed above.

TABLE 3

η	g_1	g_1'	g_1''	η	g_1	g_1'	g_1''
0·0	0·000	0·000	0·088	1·6	0·013	−0·010	−0·010
0·1	0·000	0·008	0·068	1·7	0·012	−0·010	−0·003
0·2	0·002	0·014	0·049	1·8	0·011	−0·010	0·002
0·3	0·003	0·018	0·030	1·9	0·010	−0·010	0·006
0·4	0·005	0·020	0·013	2·0	0·009	−0·009	0·009
0·5	0·007	0·020	−0·002	2·1	0·008	−0·008	0·011
0·6	0·009	0·019	−0·015	2·2	0·007	−0·007	0·012
0·7	0·011	0·017	−0·025	2·3	0·007	−0·006	0·011
0·8	0·012	0·014	−0·033	2·4	0·006	−0·005	0·011
0·9	0·014	0·011	−0·037	2·5	0·006	−0·004	0·009
1·0	0·015	0·007	−0·039	2·6	0·005	−0·003	0·008
1·1	0·015	0·003	−0·038	2·7	0·005	−0·002	0·007
1·2	0·015	−0·001	−0·035	2·8	0·005	−0·002	0·005
1·3	0·015	−0·004	−0·030	2·9	0·005	−0·001	0·004
1·4	0·014	−0·006	−0·023	3·0	0·005	−0·001	0·003
1·5	0·014	−0·008	−0·017				

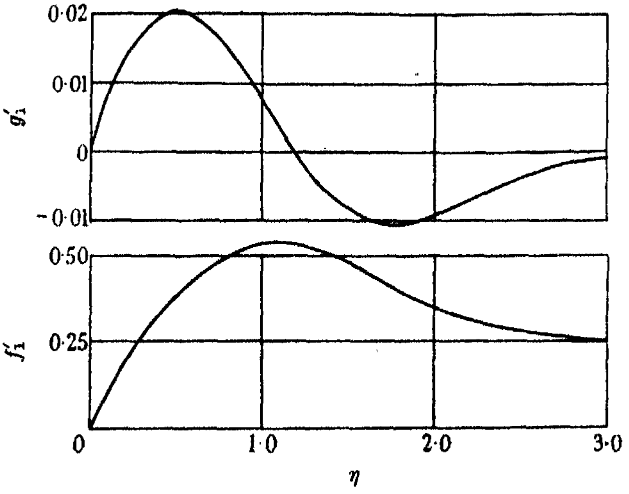


FIGURE 7

The function g_1 has been calculated and is tabulated in table 3; g_1' is shown graphically in figure 7 together with f_1' (on a different scale) for comparison; $g_1''(0)$ is equal to 0·0878. It will be seen that when $M_0 = 1$ the effect of compressibility is to increase numerically the coefficient of $8x^*$ in the velocity series in, say, the inner third of the boundary layer and to decrease it in the remainder. The maximum percentage increase is 6 at $\eta = 0·5$ and the maximum percentage decrease is 3 at $\eta = 1·8$. This then corresponds to a decrease of total velocity in the inner third and an increase in the remainder. Furthermore when $M_0 = 1$, we have from equation (123)

$$\frac{4\tau}{\rho_0 \nu_0^{\frac{1}{2}} U_0^{\frac{3}{2}}} = \left(\frac{p}{p_0}\right)^{\frac{1}{2}} [1{\cdot}3284 - (8x^*)(1{\cdot}1083) + O(8x^*)^2], \tag{124}$$

so that when $8x^* = 0{\cdot}1$ which is as far as we can safely proceed without further terms

$$\frac{4\tau}{\rho_0 \nu_0^{\frac{1}{2}} U_0^{\frac{3}{2}}} = 1{\cdot}228. \tag{125}$$

This value is substantially the same as that (1·226) obtained at the same point in the incompressible solution. The decrease in skin friction associated with the velocity corrections in the x, Y plane is in fact in this case just offset by the scale correction factor $(p/p_0)^{\frac{1}{2}}$ in passing from the (x, Y) to the (x, y) plane.

The main interest of this problem presumably lies in obtaining the point of separation and this is of course determined completely by conditions in the x, Y plane since $\partial u/\partial Y$ and $\partial u/\partial y$ vanish together. Separation is therefore given by

$$f_0''(0) - 8x^*\{f_1''(0) + M_0^2 g_1''(0)\} + (8x^*)^2 \{f_2''(0) + M_0^2 g_2''(0) + M_0^4 h_2''(0)\} + \ldots = 0. \tag{126}$$

The structure of the equations (118) and (119) introduced by compressibility considerations suggests that $g_2''(0)$ and $h_2''(0)$ will be appreciably smaller in magnitude than $g_1''(0)$ and of the same sign as $g_1''(0)$. Thus the effect of compressibility in the coefficient of $8x^*$ tends to move separation forward whilst the corresponding effect in the coefficient of $(8x^*)^2$ is to move it backwards. To see the order of the effects involved let us assume that all the compressibility terms save $g_1''(0)$ may be neglected when $M_0 = 1$. We then find, following out the same procedure as the incompressible work (Howarth 1938), that for separation x^* lies between 0·115 and 0·123 as compared with the interval 0·119 to 0·129 found for incompressible flow. It would appear then that when $M_0 = 1$ the separation point should not move forward by more than 4 % of the distance from the leading edge to the point of separation in incompressible flow.

It appears, in fact, to be impracticable to use this series method to determine the influence of compressibility on separation in view of the number of functions it would be necessary to calculate. It is, however, open to us to make use of the Kármán-Pohlhausen method which though not very accurate should be adequate to give results at least qualitatively correct.

With $U = U_0 - U_1 x$ equation (46) becomes

$$\frac{d\lambda}{dx} = -\frac{U_1}{U_0 - U_1 x}\left[g(\lambda) + \gamma M^2\{\lambda^2 h(\lambda) + \lambda\} + \frac{(\gamma - 1)}{2} M^2 j(\lambda)\right], \tag{127}$$

where, changing the notation slightly from that in § 5, M is the local Mach number corresponding to the velocity $U_0 - U_1 x$. M is therefore a known function of x and equation (127) requires a numerical or graphical integration. However, as a first approximation since the velocity change from leading edge to separation in incompressible flow is known to be relatively small (the series approach gives 12 % and

Pohlhausen's method itself gives $15\frac{1}{2}$ %) it appears plausible to replace M^2 in (127) by a mean value $\overline{M^2}$. If we adopt this approximation then

$$\log\frac{(U_0-U_1x)}{U_0}=\int_0^\lambda\frac{d\lambda}{g(\lambda)+\gamma\overline{M^2}\{\lambda^2h(\lambda)+\lambda\}+\frac{1}{2}(\gamma-1)\,\overline{M^2}j(\lambda)},\tag{128}$$

since λ is zero at the leading edge. The value x_s of x at separation is given then by

$$\log\left(1-\frac{U_1x_s}{U_0}\right)=-\int_{-12}^{0}\frac{d\lambda}{g(\lambda)+\gamma\overline{M^2}\{\lambda^2h(\lambda)+\lambda\}+\frac{1}{2}(\gamma-1)\,\overline{M^2}j(\lambda)}.\tag{129}$$

This leads to the following values for U_1x_s/U_0 when $\gamma = 1{\cdot}4$:

$\bar{M}$	U_1x_s/U
0	0·156
1	0·148
$\sqrt{10}$	0·105
10	0·044

These values of U_1x_s/U_0 imply a change in M^2 from leading edge to separation of 31 % when $\bar{M} = 1$, 43 % when $\bar{M} = \sqrt{10}$ and 66 % when $\bar{M} = 10$. It therefore appears desirable to reject this approximation and to integrate equation (127) as it stands; this has been done graphically. The results obtained are shown in the following table where M_0 refers to the Mach number at the leading edge.

M_0	U_1x_s/U_0	Mach number at separation
0	0·156	0
1	0·148	0·83
$\sqrt{10} = 3{\cdot}16$	0·107	2·4
10	0·052	5·4

The results do not differ as much as might be expected at first sight from the approximate ones obtained above if $\bar{M}$ is interpreted as the leading edge Mach number. This is due to the fact that $g(\lambda)$ is least and $k(\lambda)$ is greatest at the leading edge so that the value of M^2 plays its most important part there.

The percentage movement of separation between $M_0 = 0$ and 1 is 5 and may be compared with the outside figure of 4 % obtained from the series solution.

†The changes in the position of separation as M_0 increases beyond 1 are quite large and indicate that, for the same mainstream *velocity distribution*, as the Mach number increases the point of separation moves forward quite rapidly in the supersonic region. A somewhat similar problem has been discussed by Illingworth (unpublished) who, assuming viscosity and conductivity constant, has considered the effect of Mach number on separation for a *given linearly increasing pressure* along a flat plate. Although not strictly comparable we can consider the relation between the two sets of results by imagining the incident pressure kept constant and effecting the changes in Mach number by incident density variations. Thus in the problem of the present paper we have, if the suffix zero refers to the leading edge

$$\frac{p}{p_0}=\left[1+\frac{(\gamma-1)}{2}M_0^2\left(1-\frac{U^2}{U_0^2}\right)\right]^{\gamma/(\gamma-1)},$$

---

† The contents of this paragraph have been amended in consequence of a query raised by Professor Goldstein.

and $$\frac{1}{p_0}\frac{\partial p}{\partial x} = \gamma M_0^2 \alpha(1-\alpha x)\left\{1+\frac{(\gamma-1)}{2}M_0^2(2\alpha x-\alpha^2x^2)\right\}^{1/(\gamma-1)},$$

where $\alpha = U_1/U_0$. The considerations of the present paper can thus be applied crudely to Illingworth's problem if we make α vary with M_0 so that αM_0^2 remains constant. Illingworth's results are calculated at $M_0 = 0{\cdot}1$, $1{\cdot}0$ and $\sqrt{10}$ so that if we take $\alpha = 1$ when $M_0 = 0{\cdot}1$ we must have $\alpha = 0{\cdot}01$ when $M_0 = 1$ and $\alpha = 0{\cdot}001$ when $M_0 = \sqrt{10}$. The results tabulated above then show that separation occurs (since $\alpha = U_1/U_0$) at $x = 0{\cdot}156$ when $M_0 = 0{\cdot}1$, at $x = 14{\cdot}8$ when $M_0 = 1$ and at $x = 107$ when $M_0 = \sqrt{10}$ (assuming as is reasonable that when $M_0 = 0{\cdot}1$ separation occurs at the incompressible value $0{\cdot}156$ for U_1x/U_0). Thus the distances from leading edge to separation at these three Mach numbers are in the ratios 1 : 95 : 686 whereas Illingworth's values are in the ratio 1 : 118 : 4216. The basis of this crude comparison becomes increasingly invalid as M_0 increases but it serves to substantiate qualitatively the large backward movements of separation found by Illingworth when the Mach number is increased and the pressure gradient remains constant. On the basis of this comparison it appears to be more convenient to consider compressibility effects on separation by fixing attention on a given mainstream velocity distribution rather than on a given pressure distribution since the magnitude of such effects is thereby much reduced.

11. GENERAL DISCUSSION OF THE THEORY AND ITS RESULTS

The assumptions $\mu \propto T$ and $\sigma = 1$ have been shown to lead to a considerable simplification of the boundary layers equation in compressible flow. This simplification is achieved by a transformation

$$Y = \left(\frac{p}{p_s}\right)^{\frac{1}{2}}\int_0^y \frac{T_s}{T}\,dy \tag{130}$$

of the co-ordinate y normal to the boundary. In the (x, Y) plane the equation of motion can be transformed into

$$\frac{\partial^2\chi}{\partial x\partial Y}\frac{\partial\chi}{\partial Y}-\frac{\partial^2\chi}{\partial Y^2}\frac{\partial\chi}{\partial x} = U\frac{dU}{dx}\left[1+\frac{(\gamma-1)}{2a_1^2}\left\{U^2-\left(\frac{\partial\chi}{\partial Y}\right)^2\right\}-\frac{\gamma}{2a_1^2}\chi\frac{\partial^2\chi}{\partial Y^2}\right]+\nu_s\frac{\partial^3\chi}{\partial Y^3}, \tag{131}$$

which is identical with the standard incompressible form apart from the term

$$G \equiv 1+\frac{(\gamma-1)}{2a_1^2}\left\{U^2-\left(\frac{\partial\chi}{\partial Y}\right)^2\right\}-\frac{\gamma}{2a_1^2}\chi\frac{\partial^2\chi}{\partial Y^2} \tag{132}$$

in square brackets which is simply unity in that case. The boundary conditions are $\chi = 0$, $\partial\chi/\partial Y = 0$ when $Y = 0$, $\partial\chi/\partial Y \to U$ when $Y \to \infty$ and again are identical with the corresponding incompressible conditions. The relation between velocity and temperature is, since $\sigma = 1$,

$$Jc_pT+\tfrac{1}{2}u^2 = \text{const.} \tag{133}$$

The solution of a problem of compressible flow then falls into two parts. First of all we have to solve the flow equation (131) in the x, Y plane† and then we have to

---

† This is made possible by the form of the boundary conditions which are applied at 0 and ∞ and so are independent of the (unknown) Y scale.

transform the results into the actual x, y plane by means of the inverse transformation.

$$y = \left(\frac{p_s}{p}\right)^{\frac{1}{2}} \int_0^Y \frac{T}{T_s}\, dY, \tag{134}$$

and the energy integral (133).

In problems in which dU/dx vanishes such as flow along a flat plate and flow in wakes and jets equation (131) is identical with the incompressible equation and the co-ordinate Y can be interpreted as the distance y_i normal to the layer in the incompressible flow.

The effects of compressibility are then determined entirely by the change of scale given by equation (134). The skin friction along the flat plate is found to be independent of Mach number but the boundary layer thickness increases considerably being double the incompressible thickness at a Mach number of 3·5.

In all other problems the alterations produced by the factor G and the change of scale have both to be obtained. Two examples have been considered in detail. First, for a flow in which the velocity in the mainstream increases linearly with distance from a stagnation point a solution of (131) in series has been obtained. Up to mainstream Mach numbers of unity the solution of (131) differs but slightly from the corresponding incompressible solution. Here again then the most important changes which occur arise from the change of scale which again leads to a thickening of the boundary layer and in this problem to a reduction by a factor $(p/p_0)^{\frac{1}{2}}$ in the skin friction p being the mainstream pressure and p_0 the stagnation pressure.

Secondly, the problem of flow along a flat plate with a linearly retarding pressure gradient has been considered both by the series method and Pohlhausen's. Separation which is one of the features of such a flow is determined entirely by equation (131) since $\partial u/\partial y$ and $\partial u/\partial Y$ vanish together. The change of scale is important in determining the boundary layer thickness and the variation of skin friction along the surface up to the point of separation. It is fairly evident on general grounds that for a given retarding mainstream velocity distribution (131) will lead to a forward movement of separation with increasing Mach number. For, the effect of the pressure gradient term $U(dU/dx)$ is enhanced by a factor $G_0 = 1 + \frac{(\gamma - 1)}{2a_1^2} U^2$ at the boundary so that even though the term G may fall below unity in the outer part of the layer one would expect earlier separation.

This prediction is certainly fulfilled in the particular problem solved. Pohlhausen's method gives a forward movement of separation of 5 % at $M = 1$, 33 % at $M = 10$ and 66 % at $M = 10$ these percentages being percentages of the distance from leading edge to separation in incompressible flow.

We can consider qualitatively the general effects of compressibility in the following way. Let us fix attention on a certain mainstream velocity distribution say one which rises from zero at a stagnation point to a maximum and then falls off again. Let us imagine that this distribution is maintained unaltered and take as our standard conditions those in the mainstream at the stagnation point and denote them by a suffix zero. Then we can examine compressibility effects by considering a sequence of values of a_0; increasing compressibility will of course correspond to decreasing a_0.

Considering a particular value of a_0, the mainstream Mach number U/a_1 is given by $\dfrac{U}{\{a_0^2-\frac{1}{2}(\gamma-1)\,U^2\}^{\frac{1}{2}}}$ and the pressure ratio p/p_0 by

$$\frac{p}{p_0}=\left[1-\frac{(\gamma-1)}{2}\frac{U^2}{a_0^2}\right]^{\gamma/(\gamma-1)}.$$

The true ordinate y is determined from Y by the relation

$$y=\left(\frac{p_0}{p}\right)^{\frac{1}{2}}\int_0^Y\frac{T}{T_0}\,dY=\left(\frac{p_0}{p}\right)^{\frac{1}{2}}\left[Y-\frac{(\gamma-1)}{2a_0^2}\int_0^Y u^2\,dY\right]$$

and the skin friction τ by

$$\tau=\mu_w\left(\frac{\partial u}{\partial y}\right)_w=\mu_0\left(\frac{p}{p_0}\right)^{\frac{1}{2}}\left(\frac{\partial u}{\partial Y}\right)_0.$$

If we start by considering a_0 very large we have the ordinary incompressible solution (with kinematic viscosity $\nu_0=\mu_0/\rho_0$) which gives in particular $Y=y_i$ (the incompressible ordinate) and a distribution of skin friction which increases from zero to a maximum and then falls off to zero again if the retarding pressure gradient is sufficiently maintained.

Now consider a finite value of a_0. The velocity u in the x, Y plane will differ from the incompressible velocity u_i at the same point of x, y_i plane by an amount depending on the factor G. In the accelerated region one would expect u to increase more rapidly than u_i near the boundary since $G>1$ and then to increase more slowly than u_i in the outer part of the layer. In the retarded layer on the other hand one would expect u to increase less rapidly than u_i near the boundary and more rapidly in the outer part of the layer. Thus we should expect $(\partial u/\partial Y)_0$ to be greater than $(\partial u_i/\partial y)_0$ in the accelerated region and less in the retarded region. The factor $(p/p_0)^{\frac{1}{2}}$ will be less than or equal to unity everywhere and hence we should expect a decreased skin friction coefficient in the retarded region on both counts.† As remarked above separation is determined entirely by $(\partial u/\partial Y)_0$ and would on the whole be expected to occur earlier than in the incompressible flow. For although there is some increase in $(\partial u/\partial Y)_0$ in the accelerated region, if the two examples discussed are typical, this increase will be more than offset by the decrease in the retarded region.

In the accelerated region the two factors affecting the skin friction produce opposite effects and it is not possible to draw any general conclusions. If the accelerated flow solution in § 8 can be taken as a guide the factor $(p/p_0)^{\frac{1}{2}}$ predominates and leads here too to a reduction in skin friction coefficient.

The actual boundary layer ordinate is given by

$$y=\left(\frac{p_0}{p}\right)^{\frac{1}{2}}\left[Y-\frac{(\gamma-1)}{2a_0^2}\int_0^Y u^2\,dY\right]>\left(\frac{p_0}{p}\right)^{\frac{1}{2}}Y\left[1-\frac{(\gamma-1)}{2a_0^2}U^2\right]$$
$$\doteqdot y_i\left[1-\frac{(\gamma-1)}{2a_0^2}U^2\right]^{-(2-\gamma)/2(\gamma-1)}.$$

---

† This statement is not at variance with the results found for the flat plate with a retarding pressure gradient in § 10 for there standard conditions were defined by values at the leading edge and not stagnation values as here. Thus the incompressible solution taken as standard in the present paragraph corresponds to higher kinematic viscosity and hence to higher skin friction than that referred to in § 10.

Thus the boundary layer ordinate y is always greater than the corresponding ordinate y_i of the incompressible flow and one would therefore expect increased boundary layer thickness throughout the whole layer.

To sum up then, the effects of compressibility on such a boundary layer would appear to be

(i) a reduced skin friction coefficient everywhere,
(ii) an increased boundary layer thickness,
(iii) an earlier separation,

as compared with the corresponding problem in incompressible flow with kinematic viscosity $\nu_0 = \mu_0/\rho_0$.

REFERENCES

Brainerd, J. G. & Emmons, H. W. 1941 *J. Appl. Mech.* **8**, 105.
Brainerd, J. G. & Emmons, H. W. 1942 *J. Appl. Mech.* **9**, 1.
Cope, W. F. & Hartree, D. R. 1948 *Phil. Trans.* A, **241**, 1.
Crocco, L. 1946 *Monograf. Sci Aeronaut.* no. 3.
Howarth, L. 1934 *Rep. Memor. Aero. Res. Comm.* no. 1632.
Howarth, L. 1938 *Proc. Roy. Soc.* A, **164**, 547.
Kármán, Th. von & Tsien, H. 1938 *J. Aero. Sci.* **5**, 272.

Intensity measurements in the very soft X-ray region

By F. C. Chalklin

Canterbury University College. Christchurch, New Zealand

(*Communicated by E. N. da C. Andrade, F.R.S.—Received* 29 *November* 1947)

In the grazing incidence grating spectroscopy of the soft X-ray region, it is customary to measure the intensity distribution in the emission bands from solids by methods involving the assumption of the photographic reciprocity law. A procedure has been evolved by which this assumption is avoided. An account is given of the experimental technique and of the essential precautions. Characteristic blackening curves are obtained for the photographic plates in the soft X-ray region and under the actual conditions of spectroscopy.

Intensity distribution curves have been obtained for the K emission bands of graphite, lampblack, diamond and carborundum. The emission band of lampblack is similar to that of graphite but is somewhat wider. The difference between the present curves and those previously obtained appears most clearly in diamond and carborundum. The energy distribution of the p electrons in the valence bands of the solids is calculable from the curves.

INTRODUCTION

The discovery, by Compton & Doan, that X-ray spectra of adequate intensity and adequate dispersion could be obtained with a ruled grating in the grazing incidence position, made possible the spectroscopy of X-rays of wave-lengths greater than 20 A. With the method of calibration by the vacuum spark, and with the replacement of plane by concave gratings, the modern soft X-ray spectrographs have become instruments of high precision and of high resolution. One of the most important branches of soft X-ray spectroscopy has been the study of those radiations which result from electronic transitions in which valence electrons are involved. These radiations, from solid anti-cathodes, are not monochromatic but are spread out into bands of appreciable widths. The widths of these bands, and the intensity distribution within them, yield the energy distributions of the valence electrons of the solid radiators. Obviously, therefore, the precise examination of the bands is a matter of some importance in the study of the solid state. Two processes are involved—the measurement of wave-length and the measurement of intensity. The accuracy of the former is ample, but the intensity determinations add additional complication to an already troublesome technique. It is customary to overcome the difficulty by assuming that the reciprocity law is valid for the soft X-ray region. Exposures for different times are recorded on one and the same photographic plate, and it is assumed that the variation of blackening with intensity is equivalent to the variation of blackening (density) with time. This permits a blackening-intensity characteristic curve to be drawn for the plate and therefore permits the blackening within the spectrum to be converted to intensity. The reciprocity law holds for ordinary X-rays and it may indeed do so for soft X-rays, but it is desirable to develop a method in which no such assumption is made. This was one purpose of the experiments which are to be described in this paper. The second purpose was to examine the intensity distribution in the *K* emission bands of carbon. Siegbahn & Magnusson (1935) had examined these bands with the high resolution of the concave spectrograph. They had found them very complicated and had found that they differed very considerably for different forms of carbon.* They give precise wave-length measurements, and they give microphotometer records of their spectra, but they have not converted these records to intensities. In the present work attention is concentrated on obtaining these intensity curves and use is made of the Siegbahn & Magnusson wave-length determinations in the calibration of the plates. The experimental procedure provides a method of deriving the characteristic blackening curves of photographic plates in the actual conditions in which the spectrograms are obtained.

SPECTROGRAPH, X-RAY TUBE AND VACUUM SYSTEM

For soft X-rays the critical grazing angle is larger and is less sharply marked than for the harder radiations. Thus, in the present experiments, in which spectra at about 45 A are investigated, there is no need to use very small angles. This is shown by the reflexion experiments of Valouch (1930) for X-rays of this wave-length. The angle actually employed was 2° 53′. The use of smaller angles leads to difficulties in spectrograph design.

The grating was of glass, was concave and of radius of curvature 99·54 cm., and was ruled with 576 lines per mm. on Professor Siegbahn's Uppsala engine. The rulings were 1 cm. long and the ruled portion of the grating extended for 2 cm. The grating was mounted in a vacuum spectrograph which has been described in some detail elsewhere, Chalklin, Watts & Hillson (1938). The slit and photographic plate lay on the Rowland circle of the grating in the manner indicated in figure 1, where it is seen that the grazing angle of incidence, even when it is as large as 2° 53′, leads to an arrangement very different from that of ordinary spectroscopy. For example, the radiation strikes the photographic plate at an angle of only about 7°. As will be seen later, this setting, peculiar to soft X-ray work, adds to the technical difficulties of intensity determinations. The special feature of the spectrograph is that all the adjustments may be made by optical or mechanical means. In this way one avoids tedious trial and error focusing, which would entail, for each test, an evacuation of the spectrograph and X-ray tube. The instrument is provided with a rail turned to the radius of curvature of the Rowland circle. This rail carries the grating carriage and the plate holder and is adjusted to pass vertically under the slit. By sliding the grating carriage along the rail the angle of incidence may be changed: by sliding the plate-holder various spectral regions may be examined; and in neither case is any re-focusing necessary.

---

* The *K* emission of carbon and carbon compounds was also studied by Broili, Glocker & Kiessig (1934).

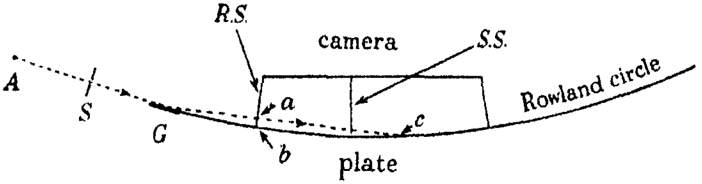


FIGURE 1. Arrangement of spectrograph in horizontal plane. Ratio slits at *R.S.*

The X-ray tube (described elsewhere, Chalklin *et al.* 1938) was of fused quartz with water cooled anti-cathode and cathode focusing system and with a helical tungsten filament. The electrodes were readily removable. The tube was sealed directly to the spectrograph. The filament current was supplied by accumulators, and the tube voltage was derived from a high tension generator giving 3600 to 5000 V. To avoid the deposition on the surface of the anti-cathode of carbon from gaseous hydrocarbons or of tungsten from the filament, it was essential to maintain an excellent vacuum in the X-ray tube. In the spectrograph itself the only requirement was that the pressure should be so low that absorption by the residual gases should be negligible. This was fortunate, since the photographic plate must certainly evolve gas, and since it was not possible to de-gas the metal work. Hence in this, as in other modern forms of vacuum spectrograph, the slit structure is built into the wall of the spectrograph, and only by way of the narrow slit can vapours enter the X-ray tube.

The evacuation of both tube and spectrograph was performed by a single Metropolitan-Vickers pump unit. This consisted of an 03 backed by an 02 oil diffusion pump, which was in turn backed by a small rotary pump. The pumping lines from the X-ray tube and the spectrograph met at the top of the 03 pump, and each line included a large metal stop-cock. At the end of an exposure these cocks could be closed and the photographic plate could be withdrawn from the spectrograph without letting air into the pumps. There was therefore no need to wait for the oil to get cold. Furthermore, the cocks ensured that, under all circumstances, the oil could be shut off from the atmosphere.

The X-ray tube and the spectrograph were connected by a tube containing a stopcock, which, when opened, equalized the pressure in the two chambers when air was being admitted to the spectrograph. A liquid air condensation trap was included in the pumping line immediately adjacent to the X-ray tube as a precaution against the entry of any vapours from the oil diffusion pumps or of any vapours that might emanate from the spectrograph, travel to the pump and then diffuse back along the X-ray tube pumping line. In experiments on the radiations of various forms of carbon, it would clearly have been unwise to risk the deposit of unwanted carbon. Apart from the discharge tube vacuum indicators which were a part of the pump unit, no vacuum gauges were included in the system. They proved quite unnecessary. To isolate the spectrograph from the vibrations of the pump unit, Tombac bellows were included in the pumping lines and proper precautions were taken to prevent the fracture of the quartz X-ray tube by the unbalanced atmospheric force of some 30 lb. wt. which was thereby introduced. Altogether, this vacuum system gave entire satisfaction.

THE METHOD OF INTENSITY MEASUREMENT

The fundamental requirement of the method is the registration on one and the same plate of two exposures of the same spectrum, the times of these exposures being the same, but the intensity of the one being a known multiple of the intensity of the other. Once this is achieved, it is possible to obtain a characteristic blackening curve for the plate. With this purpose, the two-aperture arrangement of figure 3 was screwed to the end of the camera (plate-holder) shown in figure 1. To allow the passage of all the wave-lengths corresponding to the length of the photographic plate, it was necessary for the apertures to be in the form of horizontal slits. The widths of these were adjusted so that one was a convenient multiple of the other, and they were then clamped to prevent further movement.[*] When the X-ray tube was operated, two beams entered the camera, one through the upper slit, W, and one through the lower, N, and in this way two exposures of the same spectrum were obtained. A ray diagram in a vertical plane in figure 2 shows how the radiation passed from the anti-cathode, A (on the left), through the slit, S (rather more than 1 cm. long), of the spectrograph to the grating, G, and thence through the ratio apertures to the photographic plate. The separating screen ($S.S.$), which is seen in figure 2 and is also shown in figure 1, was a strip of brass, 0·4 cm. wide and with bevelled edges, the purpose of which was to prevent the over-lapping of the two beams. If we consider a point, P_1, on the plate exposed to the one beam, and the corresponding point, P_1', at the same wave-length in the other beam, then the ratio of the areas of anti-cathode radiating to these two points is equal to the ratio of the two slit apertures. Thus, if all parts of the anti-cathode were radiating at the same rate, and if one aperture were twice the width of the other, then the intensity at a point on one trace would be twice the intensity at the corresponding point on the other.

---

[*] In figures 2 and 3 the wider slit is shown formally as the upper one. In the two series of experiments described later it was actually at the bottom.

Several details were of importance in the design of the slit apertures:

(*a*) *Distance apart.* If the apertures were too far apart, the traces would run off the sides of the plate, and if they were too close together the traces would overlap dangerously. For high orders, the suitable arrangement of the separating screen would be of importance.

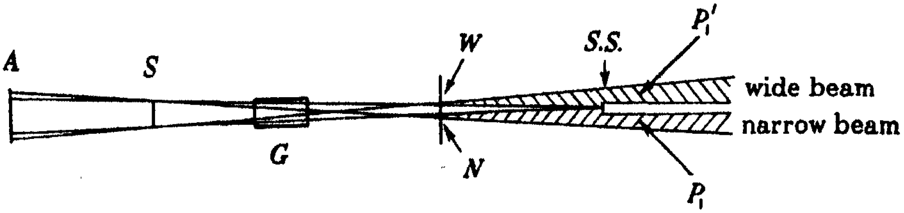


FIGURE 2. Ray diagram in vertical plane.

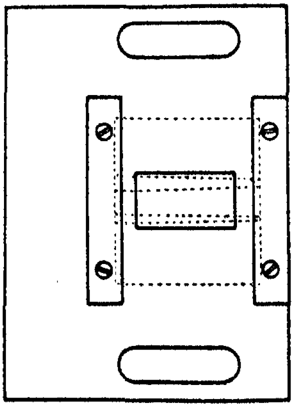


FIGURE 3. Ratio slits (movable blades dotted).

(*b*) *Width of apertures.* Too narrow apertures would render the intensity of the traces too small, whilst for too wide apertures it would be impossible to justify the assumption even of approximate radiating uniformity of the surface of the anti-cathode to which any point on the plate was exposed.

Simple geometrical calculations showed that it was possible to obtain a satisfactory compromise solution of both sets of conflicting requirements. One difficulty did, however, remain in the problem of aperture width. In this question one was merely concerned with a suitable choice for the angle subtended by the apertures at the plate. However, for a given aperture width, this angle depends on the position of the point on the plate. The angle is large for short wave-lengths and small for the longer wave-lengths. Hence the chosen aperture could not be the most satisfactory for every region of the spectrum. In fact the weak higher spectral orders would be still further weakened. It is possible to compensate for this. Inspection of figure 1 shows that different portions of the spectrum must enter the camera at different positions of the point a. The greater the wave-length, the greater is the distance bc and the greater the distance ba. Consequently, if the width of the slits was to be made variable along their length and such that their width at any point was proportional to ac, then all wave-lengths in the spectrum would be produced by the same length of anti-cathode surface. There is no point in having such exact compensation. On the other hand it is certainly desirable to prevent undue weakening of higher orders or of longer wave-lengths, and the wedge-shaped slits shown in figure 3 were therefore adopted.

In principle, the construction of the $(B-\log I)$ blackening (density) curve for a plate was a simple matter. Only relative intensities were required. A point P_1, of low blackening B_1, was chosen on the 'narrow beam' trace on the plate. An arbitrary intensity value, I_1, was ascribed to it, and it was plotted on the graph. The corresponding 'wide beam' point P_1', of blackening B_1', could then be plotted, since the intensity was $2I_1$. A 'narrow beam' point P_2 was then chosen, with blackening B_2 between B_1 and B_1', and its intensity I_2 determined from the graph by interpolation. The 'wide beam' point P_2', with blackening B_2' and intensity $2I_2$, could then be plotted. In this way the whole curve could be built up. Having obtained the form of the curve in this way, it was possible to improve the interpolations of the 'narrow beam' points, and a second, improved, curve was therefore usually plotted. When this blackening curve had been completed, it was possible to convert the blackenings in the spectrum into intensities.

So far the whole surface of the anti-cathode has been considered to have the same radiating power. This is never so. In these experiments however, by suitable arrangement of the cathode system of the X-ray tube, a large focal spot in which the radiating power diminished fairly slowly on either side of the maximum was achieved. I was concerned only with the variation in a direction parallel to the slit of the spectrograph, and this variation could be measured. The blackening of the photographic plate was determined on the peak of an X-ray line in the 'narrow' beam at a succession of points along the line. The X-ray intensities corresponding to these blackenings were computed with the aid of the characteristic blackening curve, and a graph was then plotted of intensity against distance along the line. Such a graph (figure 4) was virtually a one-dimensional pin-hole picture of the focal spot, in which the narrow aperture served as the pin-hole. The actual photometer records for the spectra (from which were taken the points P_1, P_1', etc. mentioned above) were made along tracks corresponding to the maxima of these graphs and hence corresponding to the maximum of the focal spot. In this way the portion of the anti-cathode most uniform in its radiating power was employed. Nevertheless, it was necessary to allow for variation, for it affected the ratio of the intensities of the two beams. With appropriate changes of scales, the 'pinhole' curves represented the radiating power of points along the length of the anti-cathode. Let the width of the narrow ratio slit be S, let the distance from the ratio slit to the points on the plate be d and let the distance from the ratio-slit to the anti-cathode be D. Then 1 mm. on the anti-cathode is represented by d/D mm. on the graph. Again, a point on the plate is exposed to a length of $S(D+d)/d$ on the anti-cathode. This must be represented by a length $S(D+d)/D$ on the 'pinhole' curve. If NN' in figure 4 represents this length for the value of S appropriate to the narrow ratio slit, and WW' represents the length for the wide slit, then the ratio of the mean intensity between B and B' to the mean intensity between A and A' gives the correction factor to apply to the geometrical slit ratio. Replacing the original measured slit ratio by this corrected intensity ratio, the characteristic blackening curve for the plate could be re-plotted. It may be mentioned that the largest correction in the present series of experiments amounted to 6·4 %, the usual figure being about 3 %.

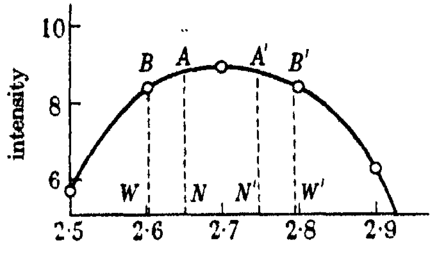


FIGURE 4. 'Pin-hole picture' of focal spot.

EXPERIMENTAL DETAILS

The photographic plates used in this work were of thin glass coated with Q2 emulsion by Messrs Ilford Ltd.* In this emulsion the usual absorption of soft radiations by the gelatin is reduced by having a very rich silver halide layer at the surface. It had already been found to be very sensitive to soft X-rays as well as being of good texture and surprisingly robust. The blackening produced by soft X-rays is, however, mainly a surface effect, and is usually small. Errors due to abrasions and to dust are therefore likely to be rather larger in the photographic photometry of the very soft X-ray region than in that of the visible spectrum. All plates were developed at approximately the same temperature, 17 to 18° C, plates L60 to L71 for 3 min. and L72 to L77 for 6 min. By using a developing dish much larger than the plate, and moving the plate about at random in it, it was hoped to avoid any variation of development over the plate surface which might be produced by the standing waves caused by ordinary rocking of the dish.

The plates were photometered on a non-recording microphotometer of simple design incorporating a barrier type photoelectric cell. The photoelectric currents were measured with a Moll short period galvanometer. Taking the photometer curves as rapidly as possible, the effects of galvanometer zero shift and change of intensity of the light beam were minimized. Correction for such small changes was made by linear interpolation. It was of great importance that the photoelectric current should be proportional to the intensity of the light falling on the cell. A number of inverse square law tests were made with various resistances in series with the cell, with various widths of the photometer slit and with two different slit-cell distances. When a light source is placed at various distances from the photometer slit, the inverse square law governs the amount of light entering the slit, but at the surface of the cell the variation is partly one of intensity and partly one of illuminated area. For this reason it would not have sufficed to make the tests with a single slit-width and slit-cell distance. Particular attention was paid to the testing of the linear relation between current and intensity because, in photometering the plates, it was found to be desirable to employ rather large resistances in series with the galvanometer and photo-cell. Such resistances were employed to prevent the over-damping of the galvanometer. They might conceivably have affected the linear relation, but the tests provided ample evidence to the contrary.

---

* Through the courtesy of the late Dr Olaf Bloch.

The photographic plates were mounted so that the edge of the exposed portion was parallel to the direction of motion of the plate carrier of the photometer, and the photometer slit was then made parallel to the spectral line under examination. The latter process was rendered difficult by the diffuseness of the soft X-ray lines. The photometer screw was then turned and the galvanometer reading was obtained for the peak of the X-ray line. The plate was then moved in its own plane in a direction parallel to the X-ray line, the displacement being read on a scale on the plate-holder, and the galvanometer was again read for the peak of the line. This was done for a number of plate displacements. The results provided the information for the 'pinhole' curves (figure 4) and permitted the final photometer records to be made at the position on the plate corresponding to the centre of the focal spot.

Since the length of the slit of the microphotometer was never greater than 7 mm. and since the magnification of the image of the X-ray line thrown upon it was 6, the effective width of plate responsible for any photometer reading was never appreciably greater than 1 mm. An increase of this length would have diminished any spurious effects due to grain size, abrasions or other imperfections on the surface of the plate, but it would have had serious disadvantages. First, it would have rendered more important the parallelism of X-ray line and slit. Secondly, since the X-ray lines were curved owing to the lateral bending of the plate in the camera of the spectrograph, an increase in the length of the slit would have diminished the 'resolution' of the photometer. Thirdly, if the slit were too long it would alter the shape of the 'pinhole curve' of figure 4, thus increasing the apparent size of the focal spot. In this way an error would be produced in the correction factor derived from the curve. The blurring of the focal spot would also be augmented by the size of the 'pinhole slit' itself. However, employing the chosen dimensions of the apparatus, use was made of a graphical construction to test the magnitude of these two blurring effects, and their influence on the correction factor was found to be negligible.

A major source of error could have arisen through faulty identification of the pairs of points P_1' and P_1 at the same wave-length on the 'wide' and 'narrow' traces respectively. They were taken from what appeared to be corresponding points on the photometer records of the carbon K band for the two traces, and the pairing was accomplished by matching the two curves as a whole. If this matching were at fault it would affect the derived blackening curve, and consequently would affect the form of the final intensity curves of the carbon bands. It was possible to check this matching, and to check the arithmetical treatment of the results, by verifying that the areas under the intensity curves of the 'wide' and 'narrow'

K bands were in the proper ratio and that the general forms of the two curves were similar.

In the actual experiments the distance from the anti-cathode to the spectrograph slit was about 5·3 cm., the distance from the slit to the grating 5 cm., from grating to ratio slits 6 cm., from ratio slits to the second order carbon bands 5 cm., and the distance from the ratio slits to the third order 7·4 cm. Only approximate values were needed for these distances. Precision was needed, however, in the determination of the widths of the ratio slits. They were measured with a travelling microscope. For plates L 61 to L 69 the wide slit had a width of 0·103$_5$ cm. at the narrow end and 0·192 cm. at the wide end, whilst the narrow slit was 0·052 cm. at the narrow and 0·095 cm. wide at the wide end. The slits were 1·4 cm. long, the third order passing through at a distance 0·45 cm. from the narrow ends (figures 1 and 3) and the second order at a distance 0·30 cm. For plates L 72 to L 76 the end widths of the wide slit were 0·099 cm. and 0·183 cm., and those of the narrow slit 0·053$_5$ cm. and 0·096$_5$ cm. From these figures a simple calculation shows that for the first series of plates the geometrical slit ratio was 2·00 for the third order and 1·99$_5$ for the second order, and for the second series 1·87 for both second and third orders. It will be noticed that although the slits were wedged shaped and roughly twice as wide at one end as at the other, they yet subtended smaller angles at the third order band than at the second, the angles for the wide slit being 0·017 and 0·024 radians respectively. Since the measurements in the carbon bands extend over a range of about 2·5 A the long-wave ends of the intensity curves are weakened by about 4 % as compared with the short-wave ends. A glance at the form of the curves shows that the effect is not important. It seemed unwise to increase the wedge angle, and, remembering that there might be some small variation in the sensitiveness of the photographic plate and in the reflecting power of the grating over the wave-length range, it appeared superfluous to apply a correction.

Before starting the actual experiments, a good deal of preliminary work was done in the attempt to obtain a long and reasonably even focal spot on the anti-cathode parallel to the slit of the spectrograph. Some of the testing was done by coating the anti-cathode with a thick layer of boron and observing the appearance of the incandescent patch on its surface when the X-ray tube was running, and some was done by preliminary exposures. In the final arrangement, the helical filament was suitably placed between, and connected to, two parallel plates. Slit, filament and focusing plates were all vertical. Figure 4 shows that at a distance of $1{\cdot}3 \times (16{\cdot}3)/(7{\cdot}4) = 2{\cdot}9$ mm. on either side of the maximum of the focal spot the radiating power has only dropped by 10 %.

Test plates. It was considered possible that there might be some systematic intrinsic ratio of intensity between the 'wide' and 'narrow' beams, quite apart from the effect of the size of the apertures. Test exposures were therefore made with apertures of equal size with the purpose of detecting any such effect and, if necessary, of allowing for it in the numerical value of the geometrical intensity ratio. The lack of equality could be computed with sufficient accuracy from the photometer records of the test exposures by employing the blackening curve of a plate in which the usual two to one slit ratio had been used.

Plates L60, L71 and L77 were such test exposures. In L60 (order III) the intrinsic difference between 'narrow' and 'wide' beams was 3 %. This was highly satisfactory, and was within the likely limit of photographic and photometric error. Evidently there was no serious intrinsic difference. In L71 (order III) the difference was about 7%. (A second order test indicated a larger difference but was discarded as unreliable owing to a blemish on the plate.) It should be mentioned that the plates between L60 and L71 show no evidence of any systematic change. The mean difference as given by L60 and L71 was thus 5 %. This error was not large and could have been a random error of the photographic photometric processes. It was to be expected that with the small densities and small contrast factors such errors would be larger than those of the visible region of the spectrum. It seemed logical to apply a correction of 5 % for this error, despite the fact that it might, at least in part, be random. The main object of this was to discover how large a change it would produce in the shapes of the curves. The intensity curves of the bands without the correction have therefore been plotted, and also, for each substance, plotted as a dotted line, the curve with the 5 % correction. The effect of the correction was extremely small. It would only have been significant in a comparison of the peak intensity with that of the background radiation. The background intensity was, however, uncertain in any case, for it must have been a mixture of orders of different wave-lengths with possibly different blackening characteristics.

The final test plate, L77, was very different. When interpreted by means of the blackening curve of L76, which was itself abnormal, it was found to show a very large difference between the beams. The difference was well outside the experimental error and showed that a change had taken place in the apparatus. It was suspected that the trouble arose through a change in form of the focal spot due to the ageing of the filament, which had already survived for a longer time than usual. Inspection of the 'pinhole' curves confirmed this, and demonstrated that the change first made its appearance in L75. Actually only two of this second series of plates have been employed, L72 quantitatively and L74 qualitatively. These points have been mentioned in detail to illustrate the importance of test exposures in this method of photometry.

Wave-length scale. An undesirable and unnecessary complication would have been introduced had vacuum spark calibrating lines been thrown on to the spectra. Hence, to obtain the wave-length scale for any of the intensity curves, the positions of the peak of the carbon band in adjacent orders were measured with a short-focus cathetometer. Next, using the wave-length of the peak as given by Siegbahn & Magnusson, a quadratic dispersion equation was derived for each plate. With this equation wave-length differences within the emission band could be calculated to an accuracy of 1 %. This is quite adequate. The wave-length scale could then be fitted to an intensity curve by identifying some point on the curve with the corresponding Siegbahn & Magnusson wave-length. In some cases this was not easy, and the absolute wave-lengths on the curves may be in doubt to about 0·1 A. Such an error would, however, introduce an uncertainty of less than 0·5% in converting wave-length differences to differences of energy and is therefore of no significance to the form of the bands.

The anti-cathodes. It is not uncommon in soft X-ray tubes for an appreciable amount of carbon to be deposited on the anti-cathodes from vapours in the tube. This could not be tolerated in the present experiments on various forms of carbon. Hence, although it was known that the vacuum conditions were extremely good and that little deposition was occurring, an exposure (L 62) was made with a copper anti-cathode to verify that no significant amount of unwanted carbon radiation was present.

Experiments were performed with carbon in the form of lampblack, 'Aquadag', Acheson graphite powder, diamond dust and carborundum powder. In every case the material was mixed with water and painted on to a thoroughly cleaned copper anti-cathode block. The coated anti-cathode was dried before being inserted into the X-ray tube. Although the surface of the anti-cathode was vertical, the materials adhered surprisingly well.

The lampblack was of commercial preparation and was stated to contain 99% of carbon. However, to reduce the possibility of any of the carbon being in hydrocarbon form, the lampblack was placed in a clean silica tube which was then evacuated and heated for 3 hr., the temperature being raised gradually to 700° C. No preliminary treatment was given to the Aquadag, the graphite powder or the carborundum powder, but the diamond dust was washed with chromic acid in case it had been handled in such a way as to leave grease on the surface of the granules.

After long preparatory pumping and running of the filament, the thermionic current in the X-ray tube was gradually raised to a normal value of rather more than 20 mA, the normal exposure amounting to about 900 mA-min. In the case of carborundum the current was kept rather lower, but, despite this, there were signs of incandescence on the anti-cathode surface. In a second experiment with carborundum (L 74) the exposure was doubled in order to obtain adequate blackening of the photographic plate in the third order. For diamond the current was only 12 mA, and the exposure 750 mA-min. Even so, there was some incandescence and some of the diamond powder was knocked from the anti-cathode, probably owing to electrostatic repulsion. Nevertheless, the appearance of the images on the plate and of the 'pinhole curve' showed satisfactory smoothness.

RESULTS

The characteristic blackening curves of the photographic plates for soft X-rays of wave-length 43 to 46 A were obtained as an intermediate step in deriving the intensity distribution in the carbon K bands. Since these blackening curves may be of some interest in themselves, two of them are shown in figures 5*a* and 5*b*. The zero of the log_{10} (*intensity*) scale is arbitrary, whilst the zero of the blackening scale in figure 5*a* refers to an unexposed portion of the plate, and in figure 5*b* refers to the absence of the plate, the unexposed portion of the plate being in this figure represented by the dotted line. The curves are similar to those for visible radiations. Their slope, small for small blackening, increases and eventually becomes constant. The values for the slope of this straight line portion of the curve vary considerably from plate to plate. Such a variation justifies the choice of a method in which the blackening curve is obtained for the actual plate on which the spectrum is registered and in which the blackening curve is actually derived from the spectrum itself. On the other hand the variation might be taken to indicate an error in the photometric technique. For this reason it is particularly gratifying that two separate plates for lampblack with different contrast factors (one having had double development) give intensity curves for the carbon band which agree well in general outline. Again, the curves for Aquadag and for Acheson graphite powder, which might be expected to agree, are seen in figure 6 to be in good general accord despite the fact that the contrast factors of the blackening curves on which they are based are respectively 0·31 and 0·25. The contrast factors are small, being usually of the order of ⅓. It might be expected that extended development would enhance the contrast factor and, there is evidence that there is at least some increase. In the second series of plates the development time was increased from three to six minutes but unfortunately only two of these plates were deemed suitable for intensity calculations. Figure 5*b* comes from one of these. The plates show no appreciable chemical fog, and in any future work it would be desirable to increase the development still further and to experiment with developers in the hope of improving the contrast and increasing the accuracy of measurement.*

It is to be pointed out that the blackening curves were obtained in rather abnormal circumstances, the radiation making an angle of only about 7° with the surface of the plate. Farineau (1938) has published a blackening curve for Gevaert X-ray film for normal incidence for the much shorter wave-length of 8·3 A. Its derivation, like that of our own curves, makes no assumption of the reciprocity relation, but is not obtained from the actual spectrograms. The curve shows the same initial increase of slope, the same straight line portion, and it also shows the familiar 'saturation' effect at large blackenings. The contrast factor is 0·85.

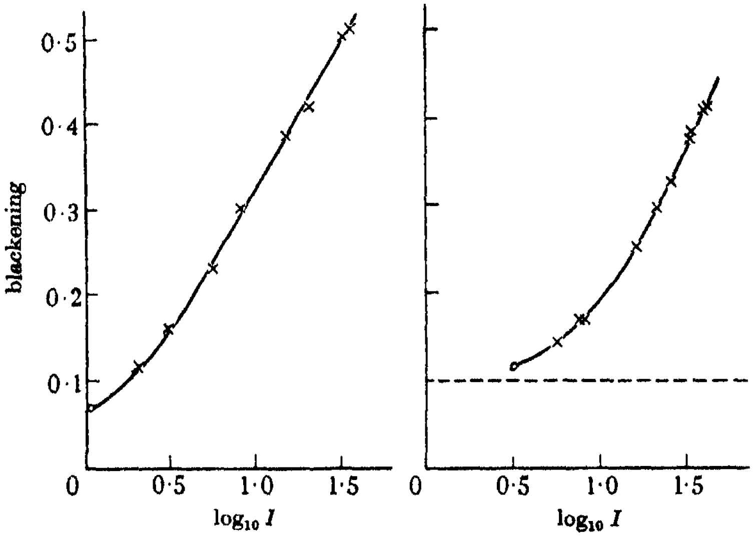
Since every successful exposure yields a 'wide beam' and a 'narrow beam' spectrum, there are at least two intensity curves for each anti-cathode examined. To make the best use of all the experimental information, the procedure of plotting mean curves has been adopted. It has the merit of diminishing errors due to dust, abrasions and grain size, which are always to be feared in photographic photometry. It has the disadvantage that it depends upon the accurate superposition of the original curves. If the superposition is inaccurate, then genuine local features of the curve may be obscured when the mean is taken. This disadvantage is perhaps superficial, for any effect lost in this way must at least have been an uncertain one. Actually the curves were superposed, not by matching any particular features on them, but by matching the effective centres of the bands. It is much the more reliable method.

Lampblack and graphite. The *K* emission band of lampblack is shown in figure 7. It is the mean of four curves, a pair from each of two plates. A check of the accuracy of the method was provided by the good agreement of the shapes of the curves derived from these separate plates. In figure 6 are shown the bands for graphite in the form of Acheson graphite powder and of Acheson Aquadag. These curves would naturally be expected to be similar, and it is satisfactory to find them so. Only in minor local details do they differ, and these differences do not appear to be significant. The mean of the two bands has therefore been taken to represent graphite and has been plotted in figure 7.

---

* It is possible that it would be better to employ the slower but more contrasty Ilford emulsion.

A comparison of the lampblack and graphite bands is of some interest. Their general similarity is obvious. However, on measuring the widths of the bands at



a *b*

FIGURE 5

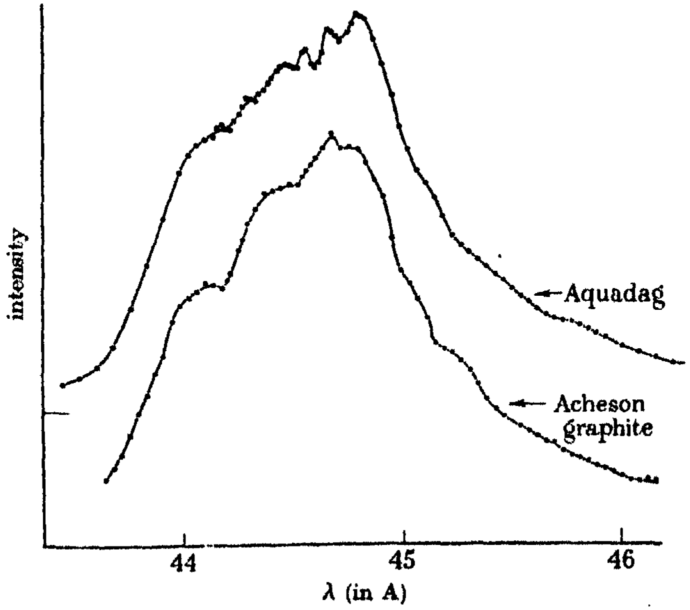


FIGURE 6

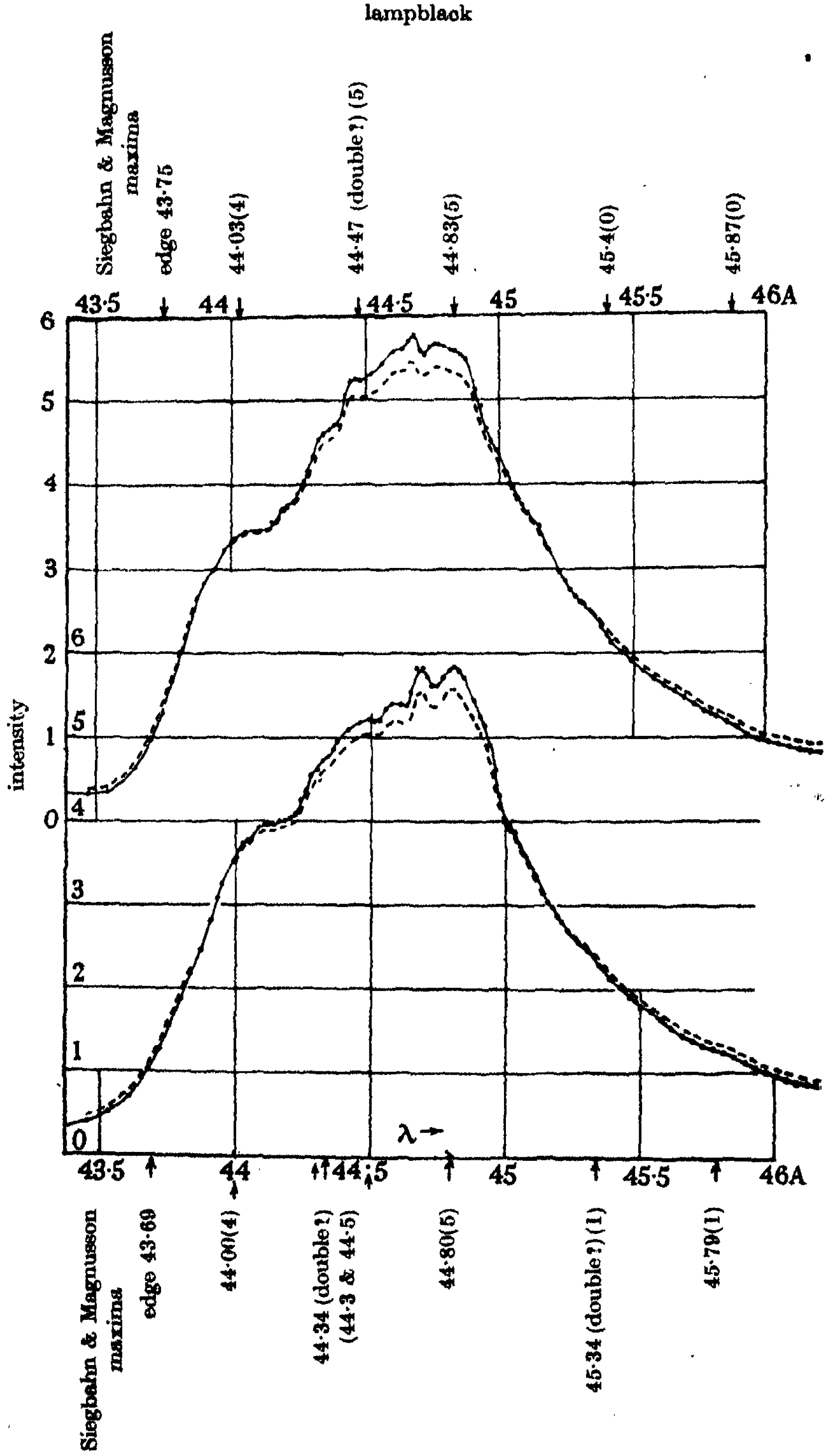


FIGURE 7. Corrected curve shown by broken line.

half maximum intensity, it was found that the value for every one of the four lampblack curves *was slightly in excess of the greatest value for the four curves* of graphite. For the mean curves and with allowance for the background radiation the values for lampblack and graphite become 1·28 and 1·22 A respectively, corresponding to 8·0 and 7·6 V. For the corrected curves the corresponding figures are 1·33 and 1·25 A, 8·3 and 7·8 V. The difference might just possibly be ascribed to error in the characteristic blackening curves. This, however, is unlikely, for the short-wave 'shoulder' of the band at about 44·1 A is actually higher for graphite than for lampblack, and if this difference in height were due to a blackening curve error, it would need to be an error in the opposite direction. The difference in width *therefore appears to be real.*

Both substances show a short-wave shoulder at about 44·1 A: both show a steep rise from 44·22 A: both show complexity between this point and the main peak; and, beyond the peak, both fall away gradually and in similar manner. The complexity near the main peak is not certain. The separation of the components is about the limit of resolution of the spectrograph. However, inspection of the individual curves embodied in the mean curve of figure 7 and of additional curves from other plates inclines one to believe that, for graphite, the complexity shown in figure 7 may be genuine. It will be noticed that the small peaks and dips shown by graphite are repeated for lampblack. The dip at 44·14 for lampblack is not visible on the graphite curve, but a close inspection of the curves of figure 6 shows possible traces of the effect. The detailed shape is not quite the same, but, bearing in mind the inevitable random differences between individual photometer curves, it is not believed that this is significant. Indeed, the differences are no greater than those between the Aquadag and Acheson graphite curves. The intensity falls sharply beyond the main peak, and beyond about 45 A the form of the curve is found to be exponential. It may be represented by the equations $I = e^{-1\cdot73(\lambda-45\cdot78)}$ for lampblack and $I = e^{-1\cdot65(\lambda-45\cdot74)}$ for graphite on the corrected curves. The figures 45·78 and 45·74 have no fundamental significance, for they are dependent on the arbitrarily chosen* scale of I. On the other hand their equality and the reasonable *agreement of the factors* 1·73 and 1·65 *demonstrate the similarity of this portion* of the lampblack and graphite curves. The exponential form is of some interest. I am not aware of any suggestion of this form in the modern electron theory of solids, and possibly other functions would fit the curve with sufficient accuracy, but it does suggest that the bands might extend for a considerable distance in the long-wave direction. In this connexion it may be recalled that L. P. Chalklin and the writer (1932, 1936), when obtaining spectra of great intensity with an instrument of low dispersion, found distinct evidence of diffuse carbon radiation at 51 A which was shown by a photometer record to be part of a 'tail' of the K emission line. Less definite 'tails' were found for the L_α lines of Fe, Co, Ni and Cu. Skinner also has reported the existence of 'tails' on the long-wave side of soft X-ray emission bands, and has discussed them, in a more recent paper (1940). Siegbahn & Magnusson (1935) record the measurement of subsidiary effects in the long-wave slope of the K bands of lampblack and graphite, their positions being indicated at the top and bottom of figure 7. In the course of the present investigation many records have been taken, and many of them show irregularities in this region. However, in the mean curves the effects, with the exception of the inflexion at 44·99 A for graphite, are weak and their location doubtful. The careful visual determinations of Siegbahn & Magnusson may be taken to provide ample evidence of the reality of the effects, but the present work does no more than demonstrate their weakness. Graphite is included in the very comprehensive study of soft X-ray emission bands by Skinner. On his curve he marks, in this region, effects at 45·33, 45·70 and 46·6 A. The general photometric procedure of Skinner and of O'Bryan & Skinner is based on the assumption of the reciprocity law. The present curves differ from theirs in emphasizing to a greater degree the intensities of the peaks. Thus, in the case of graphite, whilst there is much similarity between the curves, the ratio of the intensity of the short-wave 'shoulder' to that of the main peak is less in the present curve (0·67 for graphite and 0·60 for lampblack)* than in that of figure 13 of Skinner's paper.

---

* The peaks of the 'wide' uncorrected curves had been brought arbitrarily to approximately the same value, and the corrected curves were adjusted to match the uncorrected at half-maximum intensity.

Diamond. The results for diamond dust are shown in figures 8 and 10. Despite the difficulty of dealing with this anti-cathode material, two successful exposures were taken, one with a long time of development and one with an abnormally short time. Unfortunately the former (L 75) had to be discarded for the reason given above. Consequently the intensity scale is perhaps not quite so reliable as usual. The intensity curve presents a striking contrast to those of graphite and lampblack, the conversion to intensity having given it a form very different from that of the photometer records, and very different from the record of Siegbahn & Magnusson. As in the case of graphite, it differs from Skinner's curve in giving more emphasis to the high intensities. It emerges in a form far simpler than that of graphite or lampblack. It has a rather definite starting point at 43·66 A (figure 8) or 283·5 V. on the corrected $N_p(E)$ curve of figure 10 and its form is only slightly concave upwards before curving in to the main maximum at 44·37 A. The precise shape near the peak is doubtful and the apparent dip at 44·25 A is without significance, but a maximum at 44·60 A would, however, appear to be real.

From about 44·7 to 45 A the curve is slightly convex upwards. Beyond this there is a very definite 'hump' in the curve at 45·74 A and another, less pronounced at 45·18 A. The slope of the short-wave side of the band is greater than that of the long-wave side. It is impossible to tell from internal evidence whether or not the humps at 45·18 and 45·74 A are part of the band itself or whether they represent some form of satellite. In any case the exponential fall of graphite is missing in the present case, and any tail effect appears to be weak. If one were to assume that the two 'humps' were separate from the main band, and were to extrapolate the curve from 45 A, a result of about 45·25 A would be obtained for the end of the band. On the corrected $N_p(E)$ curve of figure 10, the corresponding point would be at 273·5 V. giving a total width of 10·0 eV. for the band.

---

* These figures are for the corrected curves but are almost identical with the uncorrected figures. The difference between 0·67 and 0·60 is partly due to the lampblack dip at 44·14 A.

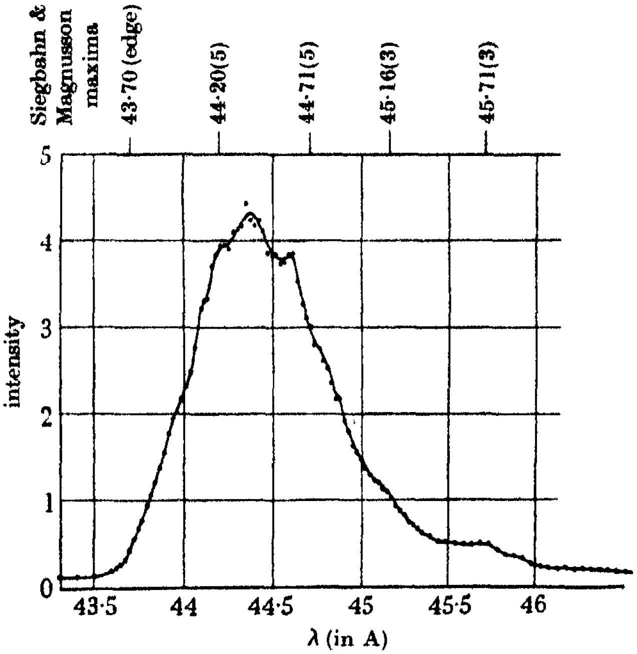


FIGURE 8. Diamond (uncorrected).

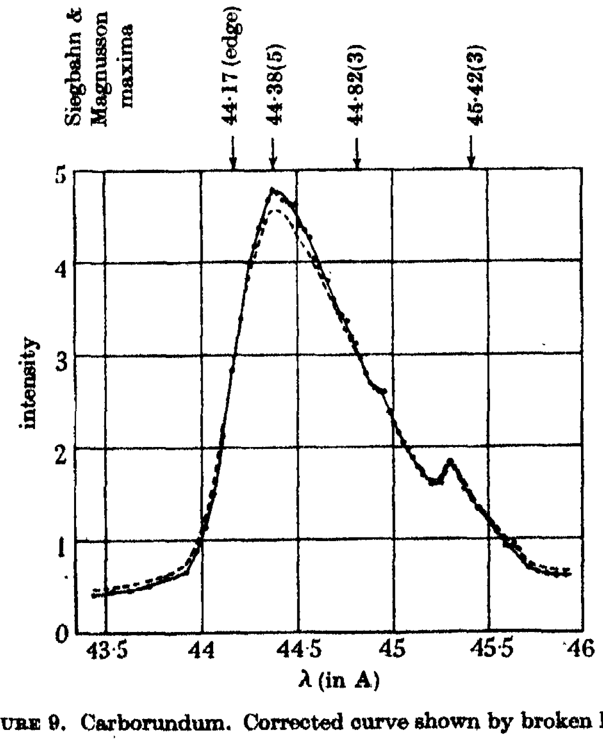


FIGURE 9. Carborundum. Corrected curve shown by broken line.

Carborundum. Of the two satisfactory plates obtained for the carbon K radiation of carborundum, one (L 67) was suitable for the photometry of the second order, and the second (L 74) was given an abnormally long exposure in order that the third order might be examined. Records and calculations were made for this plate, but, as the intensity was small and as the third order was superposed on a first order silicon $L_{\mathrm{II,\,III}}$ band of comparable intensity, the results were not quantitatively useful and are not reproduced in this paper. The second order intensity curves of L 67 were in excellent accord and their mean is shown in figure 9. Once again it differs much from the photometer records from which it was derived, and once again it emerges in a more simple form. After a start less definite than that of diamond, it rises steeply and regularly till it approaches the well-defined peak at 44·39 A. Re-inspection of the points on the blackening curve suggests that the peak might well be even sharper, and the curve on either side more nearly linear than the figure actually shows. On the long-wave side the curve descends linearly to the inflexion at 44·88 A, after which it again falls steadily till it rises to the definite sharp peak at 45·30 A. The 44·88 A inflexion, though small, is probably genuine. The shape of the curve suggests that the two 45·30 and 44·88 A effects represent subsidiary bands superposed on the main band. If this is indeed the case a rough estimate may be obtained of the width of the main band by extrapolation of the straight portions. The band would then be considered to start at 43·98 A and end at 45·36 A giving a width of 8·5 eV. On the corrected curve the figures are 43·94 A, 45·38 A and 8·9 V. Such a procedure disregards the weak but definite tail on the short-wave side of the band.

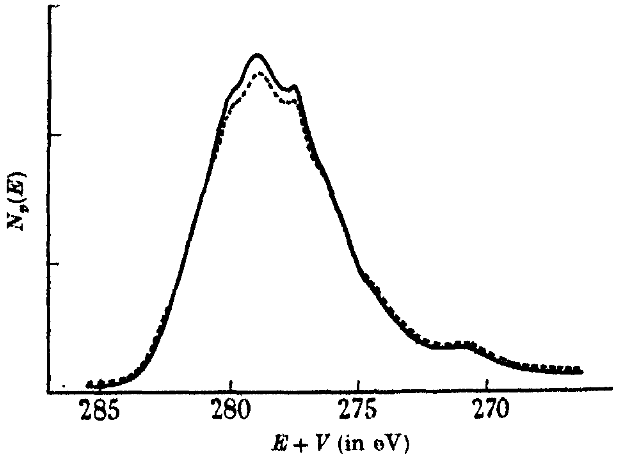


FIGURE 10. Diamond. Corrected curve shown by broken line.

Significance of the curves. Although the main purpose of this paper is the presentation of experimental results, it is desirable to give a very brief account of their relation to the theory of solids. This was investigated by Jones, Mott & Skinner (1934) and later summarized and extended by Skinner (1940). The K emission bands which we have examined are the radiations which are produced when K-ionized atoms in the crystal are refilled by valence electrons. The transitions may be described as $K-$ *valence*, and, since the K level is known to be sharp, the energy width of the emission band represents the energy range of the valence states involved in the processes. The intensity distribution within the band must depend on the energy distribution of the valence electrons, or, if the states are all occupied, the number of levels $N(E)$ per unit energy range, where the energy E of the valence levels may be measured from the lowest level. Remembering that the transition probability contains a factor ν^3, the intensity distribution $I(E)$ may be expressed as:

$$I(E) \propto \nu^3 f(E)\, N(E),$$

provided that all the levels in the range considered are occupied. At low temperatures the valence electrons occupy the lowest levels, and above a certain energy, governed by the number of electrons available, the levels are empty. At a point corresponding to this energy the emission band should come to an abrupt stop. This implies that the emission bands should have sharp edges on their short-wave side—and this has been amply verified by O'Bryan & Skinner for metals.

The periodic variation of potential through the crystal lattice causes the valence levels to be divided into bands of levels which may or may not overlap. For insulators the electrons are supposed to fill the lower bands completely and the next, empty, bands are at appreciably higher energy. Hence the application of an electric field cannot change the energy of the electrons. The high energy end of the emission band corresponds to the high energy end of the band of valence levels, and, for insulators, its form should be governed by the distribution of levels in the band. It is not expected to have a sharp edge. This is well illustrated by the results for diamond. The slope of the carborundum curve is indeed rather steep on the short-wave side, but neither this substance nor graphite, both semi-conductors, show the sharp edge which is characteristic of the partly filled bands of levels in metals.

When considering the shape of the curves it is necessary to separate the valence states of carbon into two groups—those corresponding to wave-functions which, near the nuclei, have the form of atomic s functions, and those corresponding to atomic p wave functions. Applying the usual selection rule, the expression for the intensity distribution in the K bands becomes: $I(E) \propto \nu^3 f(E)_{p\to s} N_p(E)$, where $f(E)_{p\to s}$ refers to the transitions to the K ($1s$) state from the p valence states and should in fact be constant. On this basis $N_p(E)$ is proportional to $I(E)/\nu^3$ or to I_λ/ν^5. Actually the intensity curves would represent I_λ only if the dispersion were constant over the bands. It varies by merely 3 %, and its effect is in the opposite sense to the error due to variation of angular aperture. There is no point in allowing for it. Hence, in figure 10, the values of $N_p(E)$ are obtained by dividing the observed intensities by ν^5, and by multiplying by a convenient factor so that comparison may be made with the original curve of figure 8. The abscissae, $E+V$, are the energies in eV corresponding to the wave-lengths λ. There is indeed some difference between figures 8 and 10, but it is not large. It has therefore been thought preferable to leave the curves for other anti-cathodes in their actual experimental form and to provide scales so that conversion can be made readily if desired.

Very recently the problem of the energy bands of the valence electrons of graphite has been attacked by Coulson (1947), who has published a preliminary account of his findings. Graphite consists of planes of hexagonally arranged atoms whose separation is much smaller than the distance between the planes. A group of valence electrons is allotted to σ type bonds within the planes. One electron per atom is left over. This is a π electron similar to a $2p_z$ atomic orbital with z perpendicular to the planes. Coulson finds that these π electrons form a band which is symmetrical about, and drops to zero at, its mid-point. Since the band would accommodate two electrons per atom, it is half full, and is filled in fact to the zero mid-point. The semi-conductivity of graphite is explained by the escape, at increasing temperatures, of small but increasing numbers of electrons into the upper half of the band.

Coulson's preliminary curve gives no scale and it is therefore not yet possible to compare it with my experimental results for graphite. In any case it gives a qualitative explanation of the short-wave slope of my emission band and is roughly of the same form. Coulson's curve also shows the gradually decreasing slope on the long-wave side of the band. As it stands, however, there is no explanation of the main peak of the observed emission band with its apparently complex structure. This, however, is presumably to be ascribed to transitions of the σ electrons.

To test Coulson's prediction of a half-filled symmetrical band, it will be imperative to supplement the present work by a photometric study of the carbon K absorption spectrum. The empty half of the π electronic band should obviously yield an absorption band similar to the emission band caused by the full portion. Such an absorption band was in fact found in an early soft X-ray study (Chalklin & Chalklin 1934), but, although its form certainly resembles that of the graphite emission band, we were not certain of the form of the carbon, and the width of the band, even allowing for the low resolution of the instrument employed, appears to be rather larger than that of the emission band of the present paper. It would also be interesting to seek for a weak line caused by the electrons which have escaped into the bottom of the upper half of the band. The intensity of this line should, of course, increase with rising temperature.

The difference in width at half maximum intensity between lampblack and graphite merits some discussion. Differences of temperature should have little effect on this width either on the simple Sommerfeld theory or on the basis of a completed half-band. Since, in addition, neither lampblack nor graphite is believed to have been excessively heated during the experiments, the idea of temperature as the cause of the effect may be discarded. It must be admitted that the preliminary heat treatment that was adopted for the lampblack was insufficient to guarantee the complete absence of hydrogen from the specimen. On the other hand it is extremely improbable that hydrogen was present in quantity sufficient to affect the form of the carbon band. The conclusion is that the width difference is a real difference between lampblack and graphite.

Until lately, lampblack has been considered to consist of small particles of graphite with a slightly diminished spacing in the planes of hexagonally arranged atoms, and a slightly increased spacing between these planes, together with some 'amorphous' material to account for small angle X-ray scattering. More recent X-ray evidence (Biscoe & Warren 1942) shows that carbon blacks differ from graphite in that the parallel layers of hexagons are orientated at random about their common normal. It is natural in this case that the effective spacing between the parallel planes should be increased. In fact it is about 3·47 A, as compared with 3·35 A for graphite. On the basis of this model the interatomic distance in the hexagons is found to be the same for carbon blacks as for graphite. The dimension of the parallel layer groups normal to the layers is usually less than 20 A, but both this and the dimensions of the planes themselves grow rapidly when the carbon black is heated above 1000° C.*

An increase of interatomic distance, or a decrease of electronic density, should produce a contraction of the valence bands, and hence of the emission bands. The increased spacing of the lampblack layers can therefore scarcely be responsible directly for the observed broadening. The variability of the spacing is probably the cause of the effect. Interatomic distance controls the potential energy and hence the positions of the valence bands.† Thus, in lampblack, the position of the emission band should be variable over a small range, and the observed band, made up of slightly displaced components, would be expected to show a broadening. If this explanation is correct, then the lampblack curve represents a slightly blurred graphite band and the details of its structure should be less sharp than those of graphite. It is not possible to be certain on this point, but the curves of figure 7, and in particular the form of the main peak, seem to support it.

The spectroscopic work described in this paper was performed at University College, London, in the laboratory of Professor E. N. da C. Andrade, F.R.S. It is pleasant to have the opportunity of thanking him for his constant encouragement, of thanking Mrs L. P. Chalklin for her assistance in the photometric work, and of thanking the University of London, whose grant from the Dixon Fund made possible the construction of the vacuum spectrograph.

REFERENCES

Biscoe & Warren 1942 *J. Appl. Phys.* **13**, 364.
Broili, Glocker & Kiessig 1934 *Z. Phys.* **92**, 27.
Chalklin, F. C. 1936 *Proc. Roy. Soc.* A, **155**, 366.
Chalklin, F. C. & Chalklin, L. P. 1932 *C.R. Acad. Sci., Paris*, **195**, 374.
Chalklin, F. C. & Chalklin, L. P. 1934 *Phil. Mag.* **17**, 544.
Chalklin, Watts & Hillson 1938 *Proc. Phys. Soc.* **50**, 926.
Coulson 1947 *Nature*, **159**, 265.
Farineau 1938 *Ann. Phys., Paris*, **10**, 20.
Gibson, Holohan & Riley 1946 *J. Chem. Soc.* no. 87, p. 456.
Jones, Mott & Skinner 1934 *Phys. Rev.* **45**, 379.
Mott & Jones 1936 *Theory of the properties of metals and alloys*, pp. 68 and 71. Oxford: Clarendon Press.
Siegbahn & Magnusson 1935 *Z. Phys.* **96**, 1.
Skinner 1940 *Phil. Trans.* A, **239**, 95.
Valouch 1930 *J. Phys. Radium*, **1**, 261.

---

* It should perhaps be mentioned that Gibson, Holohan & Riley (1946) have now found it possible to prepare completely amorphous carbons.

† It affects the term $-\int \phi^*(\mathbf{r})\{V-U\}\phi(\mathbf{r})\,d\tau$ of Mott & Jones (1936).

The intermediate state of superconductors

I. Magnetization of superconducting cylinders in transverse magnetic fields

By M. Désirant and D. Shoenberg

Royal Society Mond Laboratory, University of Cambridge

Communicated by Sir Lawrence Bragg, F.R.S.—Received 12 *December* 1947
—Read 3 *June* 1948)

The magnetization curves of tin and mercury cylinders were measured in transverse magnetic fields at a variety of temperatures in the superconducting range, and for various cylinder radii from $1{\cdot}8\times10^{-2}$ to $1{\cdot}7\times10^{-3}$ cm. As the cylinder radius was reduced the magnetization curve departed more and more from the simple form to be expected for an infinite cylinder of large radius, and the details of the departures agree qualitatively with the predictions of Landau's theory of the intermediate state. The experiments show also that the sharp drop of magnetization corresponding to entry into the intermediate state coincides closely with the appearance of resistance. Detailed comparison with Landau's theory (as worked out more fully by Andrew in part III) yields estimates of the surface energy at a boundary between normal and superconducting phases at the various temperatures.

Introduction

If a magnetic field is applied perpendicular to the axis of a long superconducting cylinder of circular cross-section, it is well known that on account of the zero permeability of a superconductor, the field is distorted by the cylinder, and the field is in fact twice as great at the 'equator' of a cross-section as it would be in the absence of the cylinder. It follows that at the equator the field reaches the critical value H_c for destruction of superconductivity when the applied field is only $\frac{1}{2}H_c$, and for greater applied fields Peierls (1936) and F. London (1936) have shown that the cylinder assumes a so-called intermediate state, which is a mixture of superconducting and normal phases. Experiments (de Haas, Voogd & Jonker 1934; Misener 1938) have shown, however, that resistance reappeared in the cylinder not at $\frac{1}{2}H_c$, but usually at about $0{\cdot}58H_c$.

A possible cause of this discrepancy was first suggested by Landau (1943) who made a theoretical investigation of the structure of the intermediate state and showed that owing to the necessity of creating surface energy at the boundaries between superconducting and normal phases, the intermediate state would not form until a field rather higher than $\frac{1}{2}H_c$ had been applied. Landau's theory predicts that this effect should become more marked as the cylinder diameter is reduced, and since most of the existing experimental data refer to cylinders of much the same diameter (about 0·2 mm.), a new experimental investigation seemed worth while. At the same time Landau's paper, though indicating clearly the nature of the theory, does not work it out in sufficient detail for comparison with the new experimental results, so a more detailed theoretical investigation has also proved necessary.

All previous experiments on small-diameter cylinders have been limited to resistance measurements, though in fact much useful information can be obtained from measurements of magnetization, which give a direct indication of the relative proportions of superconducting and normal phases. The present paper (I) describes the magnetization curves for thin tin and mercury cylinders at various temperatures and for various cylinder diameters; the results agree qualitatively with the theory and yield an estimate of the surface energy as a function of temperature. The second paper of this series, by E. R. Andrew (referred to as II), describes new resistance measurements which were carried out partly to clear up the contradictions of existing data and partly to investigate the correlation of resistance and magnetization curves; in particular one experiment was made in which resistance and magnetization measurements were simultaneously taken on the same specimen, and which proved useful in the interpretation of the other resistance measurements. The third paper of the series, also by E. R. Andrew (referred to as III), gives a detailed account of the developments of Landau's theory necessary for the discussion of the results of I and II. A preliminary account of some of the data in I and II was given at the Physical Society Cambridge Conference in 1946 (Shoenberg 1947; Andrew 1947), but some of the conclusions tentatively put forward then have been considerably modified.

EXPERIMENTAL DETAILS

(*a*) *Method of measurement*

The method of measuring magnetic moments has already been described (Shoenberg 1947; Désirant & Shoenberg 1947, 1948); the sensitivity was such that the magnetization curves of superconducting specimens of volume from $0{\cdot}5 \times 10^{-3}$ to $0{\cdot}5 \times 10^{-2}$ c.c. could be accurately determined. A small correction for the magnetization of the specimen mounting had to be applied; this was always estimated by measurements at fields higher than H_c and never exceeded a few per cent of the maximum magnetization of the superconductor. As will be explained below, the value of H_c cannot be deduced from the magnetization curves with the cylinder transverse to the field, and since it is important for the detailed comparison with theory, special measurements were made for some specimens with the cylinder parallel to the field. All temperatures quoted are deduced from the vapour pressure of the helium bath by means of the 1932 Leiden scale; the experimental error is unlikely to be greater than 0·005° K, though the scale itself may be in error by a greater amount.

(*b*) *The specimens*

The superconductors studied were spectroscopically pure mercury and tin (see table 1). For large diameters a single cylinder was used; the apparatus limited the permissible length to only about 8 mm. (when the cylinder was transverse to the field), and so the diameter could not exceed about 0·5 mm. without causing appreciable departure from the ideal case of an infinite cylinder. For smaller diameters two methods were used to obtain sufficient volume: (i) a row of cylinders was prepared spaced in such a way that each cylinder had practically no magnetic effect on its neighbours, and (ii) a spiral of wire was wound with the turns suitably spaced so that they had practically no magnetic effect on each other, and of a pitch such that the angle between the wire and the field did not differ too much from 90°. For one particular diameter, both such methods of construction were used (H 9 and H 2), and the experimental results showed that both kinds of specimen had similar magnetic properties.

Nearly all the cylinders and spirals were prepared by filling pyrex glass capillaries with liquid metal; one of the tin specimens (S 9), however, was made from extruded wire. In order to avoid large mechanical strains the specimens were always slowly cooled to liquid nitrogen temperature before the final cooling to liquid helium temperature and no adhesive was used in the mounting of bare tin specimens except at the ends. The capillary method cannot easily be extended to indefinitely small diameters, since for diameters less than 2×10^{-3} cm. a pressure greater than atmospheric is required in filling to overcome the pressure due to surface tension. Details of all the specimens used are given, together with the experimental results, in table 1.

EXPERIMENTAL RESULTS

(*a*) *Tin*

The transition temperature of tin is about 3·72° K, and transverse magnetization curves were taken both with increasing and decreasing fields at about 3·49, 3·00 and 2·13° K for each specimen. In a few cases measurements were made also at 3·60° K, but it was found that the slight spread of transition temperature, due to slight inhomogeneities, caused 'tails' to the magnetization curves, which above 3·5° K were sufficiently pronounced to complicate the interpretation. Typical results are illustrated in figures 1 to 6 which show the magnetization curves for one specimen (S 9) at all three temperatures, and for all the specimens at one temperature (3·00° K); one of the curves (figure 1) is shown on a larger scale to bring out certain features more clearly. In each case the ordinate is -4π times the magnetization in gauss and the abscissa the applied field h in gauss. As will be explained below, the initial slope of the graph must always be close to 2 and this fact has been used to determine the magnetization in absolute units; it has also been used to make a more precise estimate of the specimen volume (and hence radius) than is possible by more direct means. In each diagram a theoretical curve has been fitted to the experimental points by a procedure which will be explained later.

In the case of S 9, measurements of resistance were simultaneously made by E. R. Andrew, and various features of these measurements are marked in the diagram (figure 1) which will be mentioned briefly below and discussed more fully in II. The results for the longitudinal magnetization curves (field parallel to the cylinder axis) are sufficiently illustrated by figure 7 plotted in the same way as the transverse curves, and figure 8 in which the ratio of -4π times magnetization to field is plotted against field; the method of deducing the critical field H_c is discussed later (p. 74). The appropriate values of H_c are marked in all the curves.

(*b*) *Mercury*

The transition temperature of mercury is about 4·17° K, and transverse magnetization curves were taken at about 4·05, 3·97, 3·69 and 2·13° K for each specimen.

As for tin, no useful results could be obtained closer to the transition temperature. The transverse results are illustrated by figures 9 to 15 in the same way as those for tin, figure 9 being shown on a larger scale to illustrate the various features used in the interpretation of the measurements. Longitudinal measurements were made only with H 7 and are not illustrated, since they are similar in character to those for tin.

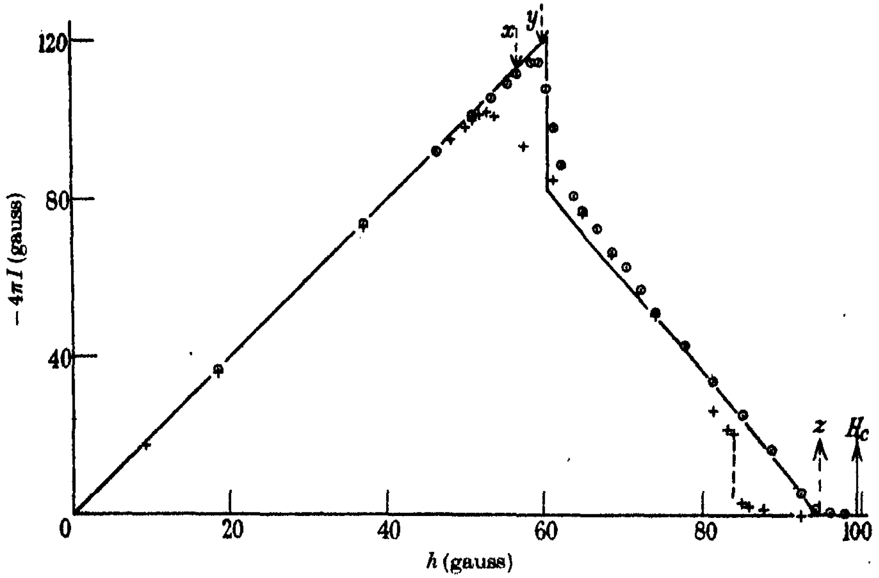


FIGURE 1. Transverse magnetization curve of S 9 at 3·000° K. ⊙ increasing fields; + decreasing fields. Full curve is theoretical assuming $\Delta'/r = 2\cdot0 \times 10^{-5}$. x, first appearance of resistance; y, rapid rise of resistance; z, full resistance restored. ------ discontinuous rise of $-4\pi I$ accompanied by discontinuous fall of resistance.

DISCUSSION OF RESULTS

The results are most conveniently discussed in two stages: first (a) pointing out the general features which are in qualitative agreement with the theory and also other features whose interpretation is more speculative, and secondly (b) showing how the results may be fitted in more detail to the theoretical predictions, thus determining the value of the surface energy between superconducting and normal phases.

(a) *General features*

For a cylinder of large diameter (where surface energy effects are negligible) we should expect to find a transverse magnetization curve of the isosceles triangle form shown in figure 16 (see, for instance, Shoenberg 1938, Chapter III), which we shall refer to as the 'bulk' curve. As shown in III, however, if the cylinder diameter is small, the theory predicts a somewhat modified curve, such as illustrated in figures 1 to 6 and 9 to 15, and shown by a broken line in figure 16, and we shall first indicate how far the experimental results agree with the theoretical predictions as regards points of departure from the 'bulk' curve.

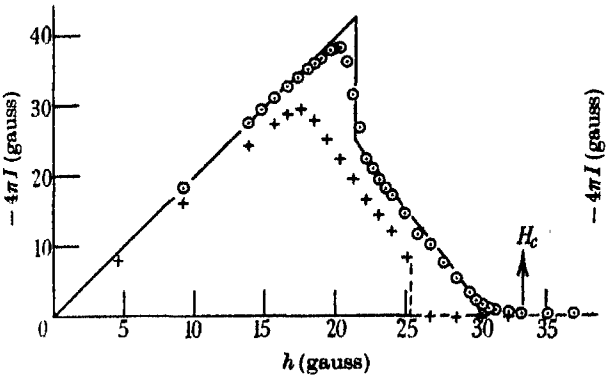


FIGURE 2. S 9, 3·490° K, $\Delta'/r = 5{\cdot}0 \times 10^{-3}$.

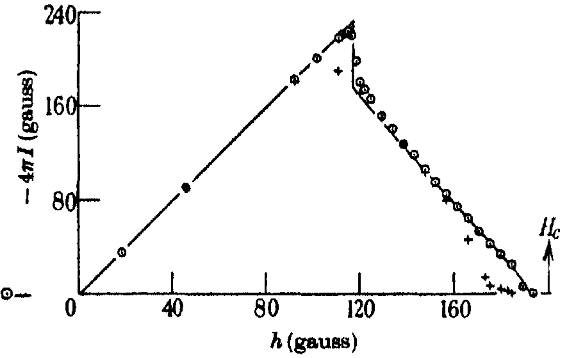


FIGURE 3. S 9, 2·121° K, $\Delta'/r = 1{\cdot}1 \times 10^{-3}$.

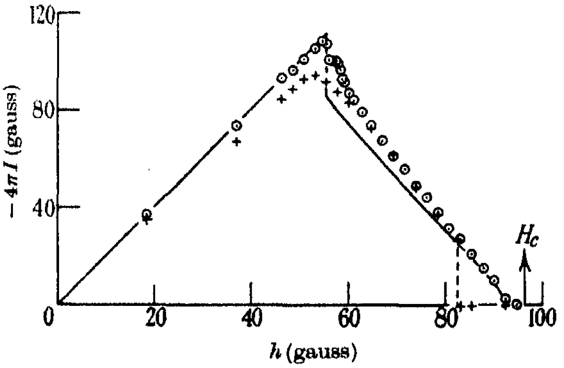


FIGURE 4. S 8, 3·010° K, $\Delta'/r = 0{\cdot}84 \times 10^{-3}$.

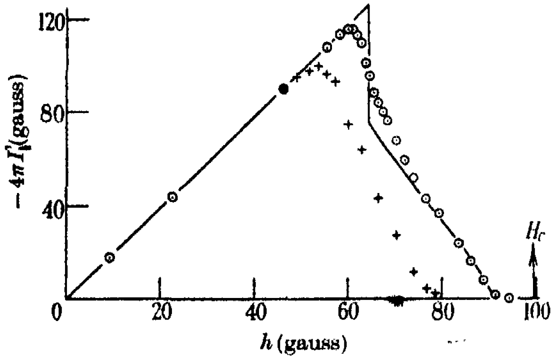


FIGURE 5. S 7, 3·006° K, $\Delta'/r = 4{\cdot}7 \times 10^{-3}$.

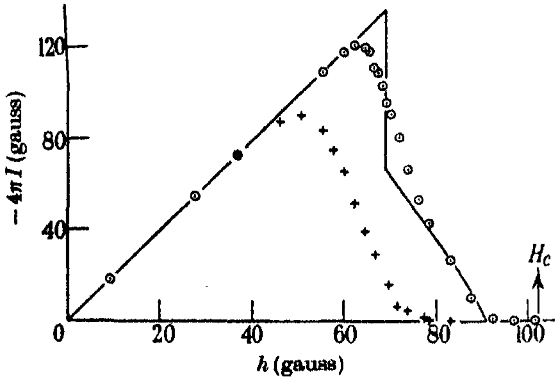


FIGURE 6. S 5, 3·002° K, $\Delta'/r = 8{\cdot}5 \times 10^{-3}$.

Figures 2 to 6 are transverse magnetization curves of various tin specimens. In each case the full curve is theoretical assuming the values of Δ'/r indicated. ------ discontinuous rise of $-4\pi I$. ⊙ increasing fields; + decreasing fields.

The main features are:

(i) Just as the theory predicts, the rising linear part of the magnetization curve continues well above $\frac{1}{2}H_c$. In this region the cylinder is wholly superconducting, and as shown in figure 1 resistance appears only when there is an appreciable departure from linearity and very close to the point at which the magnetization drops steeply.

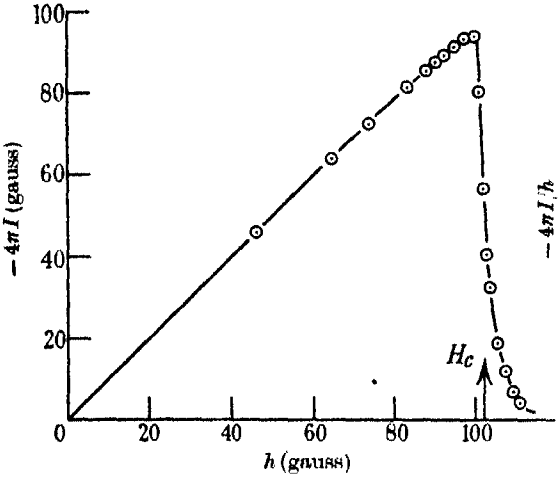


FIGURE 7. Longitudinal magnetization curve of S5 at 3·006° K.

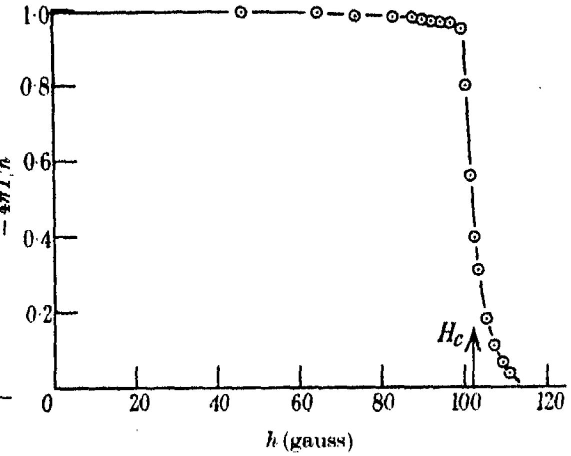


FIGURE 8. Curve of $-4\pi I/h$ against h for S5 at 3·006° K. The critical field H_c is at ordinate 0·5.

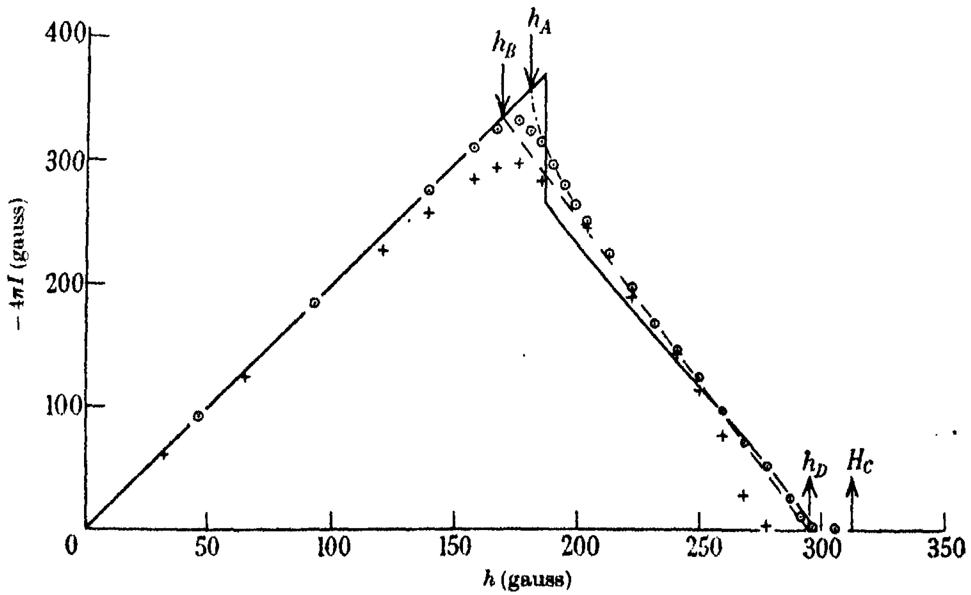


FIGURE 9. Transverse magnetization curve of H9 at 2·125° K. ⊙ increasing fields; + decreasing fields. Full curve is theoretical assuming $\Delta'/r = 1{\cdot}8\times10^{-3}$. ----- tangent to experimental curve at $B = \frac{1}{2}H_c$ to illustrate how h_B and h_D are obtained. ········ illustrating extrapolation of experimental curve to obtain h_A.

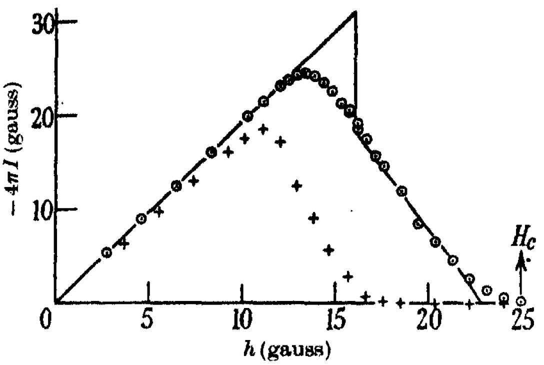


FIGURE 10. H 9, 4·046° K, $\Delta'/r = 5{\cdot}4 \times 10^{-3}$.

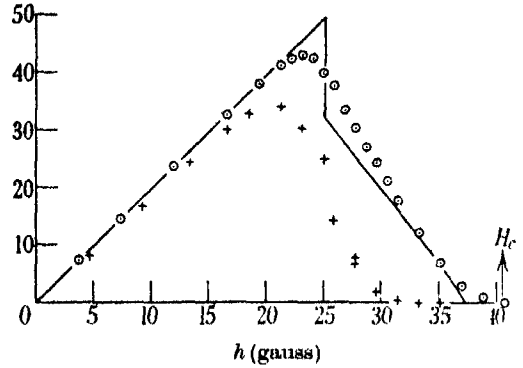


FIGURE 11. H 9, 3·962° K, $\Delta'/r = 4{\cdot}3 \times 10^{-3}$.

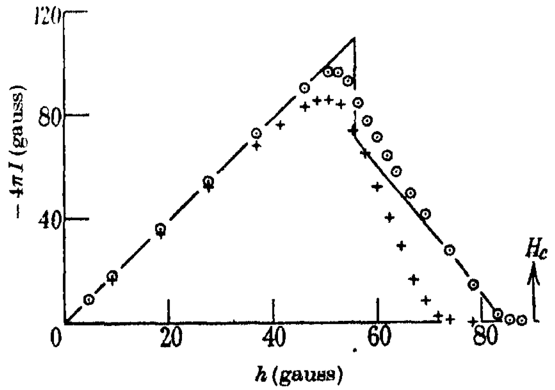


FIGURE 12. H 9, 3·691° K, $\Delta'/r = 3{\cdot}3 \times 10^{-3}$.

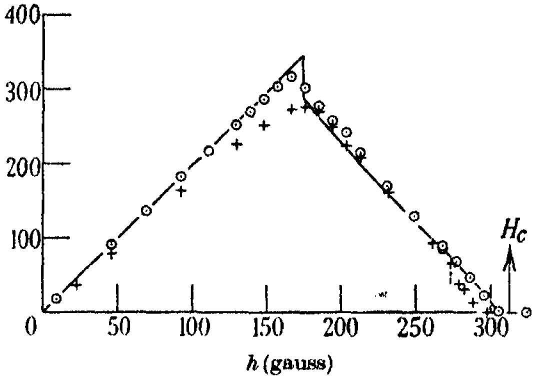


FIGURE 13. H 7, 2·115° K, $\Delta'/r = 0{\cdot}4 \times 10^{-3}$.

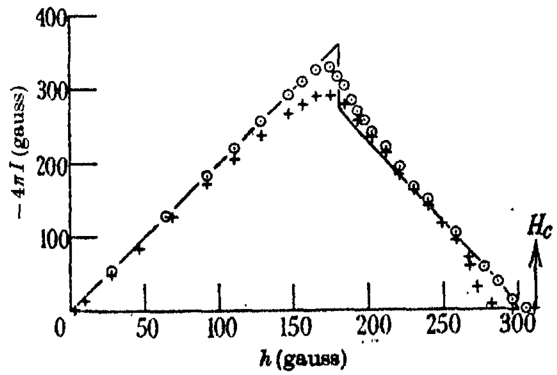


FIGURE 14. H 8, 2·121° K, $\Delta'/r = 1{\cdot}0 \times 10^{-3}$.

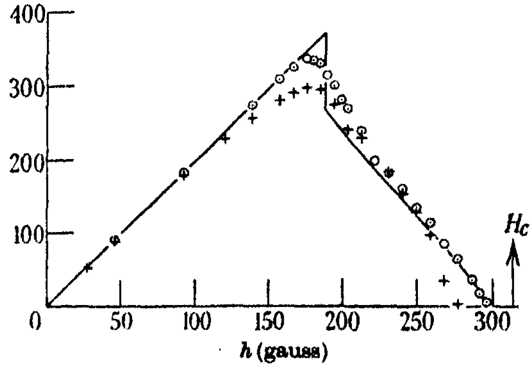


FIGURE 15. H 2, 2·120° K, $\Delta'/r = 1{\cdot}7 \times 10^{-3}$.

Figures 10 to 15 are transverse magnetization curves of various mercury specimens. In each case the full curve is theoretical assuming the value of Δ'/r indicated. ------ discontinuous rise of $-4\pi I$. ⊙, increasing fields; +, decreasing fields.

(ii) For most of the specimens there is a fairly definite field h_A at which the magnetization begins to drop steeply, and the drop is nearly vertical in the case of the extruded tin wire (S 9). The ratio $\rho = h_A/H_c$ is considerably greater than 0·5 and increases as the diameter is reduced for a given temperature, and as the temperature is increased for a given diameter. These facts are in qualitative agreement with the theory if it is supposed that the surface energy increases with temperature. The experimental values of ρ (making a somewhat arbitrary extrapolation of the kind shown in figure 9 when the experimental curve is rounded at the turn-over point) are collected in table 1. These values are not very accurate, especially for mercury, both owing to the uncertainty of the extrapolation and of the H_c values (see p. 74). Our values of ρ agree only roughly with those obtained by Andrew from resistance measurements (II) and in particular Andrew finds less temperature dependence of ρ than we do. Detailed examination of the results for S 9, where both kinds of measurements were made simultaneously, shows, however, that the values of h_A agree very well with Andrew's h_2 values and that differences arise mainly in the assumed values of H_c.

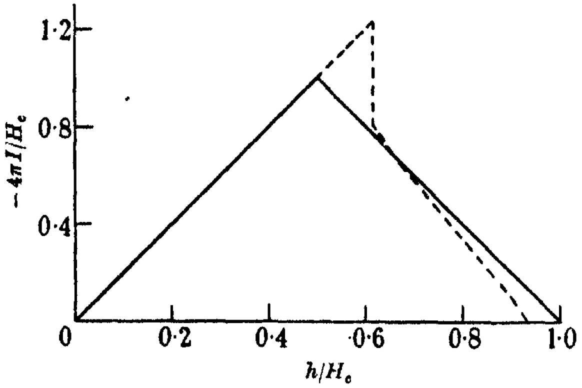


FIGURE 16. Theoretical magnetization curves. Full curve for r infinite ('bulk curve'); broken curve for $\Delta'/r = 3 \times 10^{-3}$.

(iii) The steep fall of magnetization is followed by a more gradual fall which is nearly linear, so that the initial drop somewhat resembles a horn on the magnetization curve. As will be seen from figure 16, this feature is predicted by theory, though the 'horn' is much less marked in the mercury curves than in those for tin. The cause of the difference between tin and mercury in this respect will be discussed later (see p. 78).

(iv) As already mentioned, the descending portion of the magnetization curve usually has a slight tail before the magnetization vanishes. The tail becomes relatively more pronounced as the transition temperature is approached, and since its absolute width is roughly the same at all temperatures it is probably due to a slight spread of transition temperatures within the specimen. The same observations apply also to the longitudinal curves. The width of the tail depends on the specimen, but a typical width is 3 gauss, corresponding to a spread of order 0·02° K in transition

TABLE 1

specimen	description*	r (cm. × 10^4)	ρ_{bulk}	T_c (°K)	T (°K)	H_c (gauss)	h_D (gauss)	h_A (gauss)	h_B (gauss)	ρ	σ
S 8	1 Pyrex capillary about 5·9 mg., J. & M. 12,966	179	0·502	3·711	3·490	31·6	30·6	18·5	17·2	0·583	0·560
					3·010	96·5	92·8	55·3	51·8	0·571	0·556
					2·145	197·7	195·0	110·4	104·4	0·556	0·534
S 9	spiral of extruded wire bifilarly wound about 37 mg., H. 10,000	75	0·504	3·722	3·490	33·4	30·7	20·4	17·4	0·607	0·563
					3·000	99·4	94·0	59·1	53·1	0·590	0·561
					2·121	200·5	193·4	116·0	104·2	0·573	0·535
S 7	62 Pyrex capillaries in 2 layers about 10·4 mg., J. & M. 12,966	31·6	0·512	3·724	3·489	34·0	29·7	22·2	20·1	0·642	0·664
					3·006	99·5	90·2	62·0	57·0	0·610	0·617
					2·147	199·1	184·5	120·1	110·9	0·589	0·586
S 5	209 Pyrex capillaries in 5 layers about 11 mg., J. & M. 12,966	17·3	0·504	3·738	3·490	36·2	30·2	24·1	22·6	0·661	0·743
					3·002	102·4	90·3	65·1	58·9	0·631	0·647
					2·154	200·7	181·1	126·6	110·9	0·625	0·608
H 7	6 Pyrex capillaries in 1 layer about 20 mg., H. 11,023	101·5	0·505	4·171 (4·168)	4·045	24·7 (24·0)	22·5	12·9	12·7	0·519	0·557
					3·963	40·2 (39·5)	38·0	22·9	21·4	0·565	0·559
					3·690	90·0 (89·4)	86·6	51·9	47·6	0·571	0·545
					2·115	311·9 (311·5)	307·3	175·1	161·7	0·556	0·522
H 8	32 Pyrex capillaries in 2 layers about 13·5 mg., H. 11,023	37·8	0·503	4·172 (4·165)	4·045	24·9 (23·6)	21·9	13·3	12·6	0·531	0·570
					3·963	40·4 (39·1)	36·6	23·0	21·9	0·566	0·595
					3·689	90·4 (89·2)	84·3	52·7	48·3	0·580	0·569
					2·121	312·3 (311·7)	300·3	179·3	165·9	0·571	0·570
H 9	100 Pyrex capillaries in 4 layers about 12 mg. H. 11,023	21·0	0·501	4·179 (4·172)	4·046	26·1 (24·9)	22·6	14·4	13·5	0·550	0·595
					3·962	41·7 (40·5)	37·0	24·8	23·0	0·593	0·622
					3·691	91·2 (90·1)	83·4	54·3	49·9	0·594	0·597
					2·125	312·8 (312·1)	295·2	180·2	168·6	0·575	0·570
H 2	Pyrex capillary spiral about 6·4 mg., H. 10,763	22·8	0·504	4·184	4·045	27·3	24·9	16·4	15·4	0·598	0·616
					3·963	42·4	39·5	26·0	24·2	0·607	0·609
					3·692	91·8	85·2	55·4	52·2	0·598	0·608
					2·120	313·8	302·1	182·5	169·1	0·577	0·556

* J. & M. = Johnson & Matthey, H = Hilger.

temperatures. The specimen S 8, which was a single cylinder, had about three times smaller spread, suggesting that part of the larger spread in other specimens is due to differences between one cylinder and another, rather than to variations within a single cylinder. If the tail is ignored by extrapolating the almost linear portion of the descending curve to meet the axis at a field h_D* (see figure 9), it is found that h_D is always appreciably less than the critical field H_c deduced from longitudinal measurements, the ratio h_D/H_c becoming smaller as the diameter is reduced or the temperature raised. These facts are again in qualitative agreement with theory. A useful characteristic for comparison with theory is obtained by extrapolating the almost linear descending portion (strictly speaking, the tangent to the curve at the point indicated in the footnote) backwards to meet the superconducting part of the curve at field h_B (see figure 9); the ratio $\sigma = h_B/h_D$ has a simple explicit form in the theory (see equation (4) below) and the observed values of σ for the various specimens, together with the values of h_B and h_D are also given in table 1.

(v) So far we have discussed only the 'rising field' curves, when the field has been raised monotonically from zero to above H_c. We now deal with the hysteresis features which arise when the field is reduced monotonically from above H_c to zero. In its initial stages the hysteresis seems to have the characteristics of a 'supercooling' (see, for instance, Shoenberg 1940), for in the case of a single cylinder specimen (S 8, see figure 4) the magnetization rises quite discontinuously at a field which is only roughly reproducible in repeated cycles. The discontinuity is shown as a sudden kick of the galvanometer if the field is very gradually lowered in the appropriate region. Where the specimen consists of several cylinders, separate discontinuities occur; these could be followed in detail for H 7 and the number of discontinuities was approximately the same as the number of cylinders,† but when the number of cylinders is large it becomes difficult to separate discontinuities reliably and the curve appears to be continuous. In general the degree of supercooling is greatest for small diameters and high temperatures, but the effect varies a good deal even for similar specimens. For instance, with the mercury specimens the degree of supercooling was sometimes quite different if the mercury was melted between two cycles of magnetization.

When the field is further reduced beyond the discontinuity, the return curve often almost coincides with the original descending portion, but departs from it under the horn until it rejoins the rising field curve in the superconducting region with no further hysteresis. It is probable that this behaviour (e.g. as in S 9) is characteristic of ideal conditions, but for specimens with great supercooling the descending portion of the original curve is never reached and the whole specimen becomes superconducting only at a low field (e.g. as in S 5). A further complication is that often the return curve never quite rejoins the original curve, even in the superconducting region, and a small 'frozen-in' moment is left in zero field. This frozen-in moment is always small and roughly proportional to the critical field, so it is probably due to some geometrical defect in the cylinders (e.g. sharp ends) rather than to impurity (see Shoenberg 1937). It was noticed that the frozen-in moment was smaller if a field greater than H_c was switched off suddenly rather than gradually reduced, which suggests that it may be partly associated with the way in which the supercooling is overcome. Whenever any appreciable frozen-in moment occurred the specimen was warmed above its transition temperature and recooled before the next magnetization curve was measured.

---

* For reasons which are explained in III this extrapolation has actually been done by drawing a tangent to the curve at a point where $-4\pi I = 2h - H_c$, ($B = \frac{1}{2}H_c$), in cases where the descending portion of the curve is appreciably curved.

† It is not possible to illustrate this feature clearly in figure 13, but it is quite evident at the higher temperatures.

Of the various hysteresis features the one most directly relevant to the theory is the fact that the return curve never retraces the steeply falling part of the rising field curve. This is consistent with the theoretical interpretation of the steeply falling part of the curve as an almost discontinuous transition from the superconducting to the intermediate state. We may suppose that when the field is reduced in this region the superconducting state is thermodynamically more stable than the intermediate state, but cannot be achieved until the field is further reduced because the normal regions are unable to disappear suddenly and must be 'squeezed out'. Similarly the theory suggests that in the last part of the intermediate state region, the normal state should become thermodynamically more stable than the intermediate state, but a discontinuous transition cannot be achieved because of the difficulty of squeezing out the superconducting regions; on reducing the field again, however, the normal state persists at least until the intermediate state is the more stable. Even then, however, a nucleus of superconducting phase is required for the transition to start, and so some supercooling can occur below the point where normal and intermediate states have equal free energies. The reproducibility of the horn part of the curve suggests that no analogous 'superheating' occurs in the transition from the superconducting to the intermediate state; this point is discussed further in III.

(b) Detailed comparison with theory

Two small corrections to allow for the experimental conditions have to be considered before the theory can be fitted to the experimental magnetization curves. First, our specimens are not exactly equivalent to the ideal infinite cylinder with which the theory deals, both on account of their finite length and of the mutual influence of the various cylinders within a specimen. A fair approximation is to treat the specimen as if it had a demagnetizing coefficient slightly less than the value 2π characteristic of an infinite cylinder. The difference is easily estimated from the known geometry and is always small; its effect is to shear the magnetization curve slightly in the direction of higher fields, so that even if the specimen was scaled up to make surface energy effects negligible, the value of ρ would not be exactly 0·5 but slightly higher. If this higher value is denoted by ρ_{bulk}, it is a good enough approximation to suppose that the theoretically predicted magnetization curve for the appropriate cylinder radius is also sheared to the same extent, so that the theoretical values of ρ or σ, for instance, become multiplied by $2\rho_{\text{bulk}}$. Conversely the experimentally determined values of ρ or σ should be divided by $2\rho_{\text{bulk}}$ to give the values of ρ or σ to be compared with theory; this correction has been applied in table 1. In working out the correct scale for $-4\pi I$ in the experimental graphs (figures 1 to 6, 9 to 15) we have also made an allowance for the fact that the penetration depth is not negligible; this means that even if ρ_{bulk} were exactly $\frac{1}{2}$, the initial linear rise would have a slope slightly less than 2. This effect amounts to about 1 or 2 % for the thinnest specimens at the highest temperatures.

The second correction is usually very small and arises from the small magnetic field of the current in the coil used to compensate the magnetic moment of the specimen. This field is at first proportional to the main applied field and then becomes progressively weaker as the magnetic moment falls. Since the effect of this field is again to cause a slight shear of the magnetization curve (plotted against the main applied field alone), its effect is equivalent to a slight change in the effective demagnetizing coefficient of the specimen. It is most simply dealt with by a slight adjustment in the values of ρ_{bulk} deduced from the geometry of the specimen, and this adjustment has been made in the values of ρ_{bulk} given in table 1.

In fitting theoretical curves to the experimental data it is essential to know the value of the critical field H_c. Since H_c at a given temperature varies slightly from one specimen to another, the use of published values by other authors is not accurate enough. In the case of tin, longitudinal measurements were therefore taken with the same specimens (in the case of S 9 some samples from the same piece of extruded wire were used). The value of H_c was determined as the field at which half the volume had returned to the normal state, i.e. the field at which the ratio of 4π times the magnetization to field in figure 8, had dropped to half the superconducting value. It was found that to a sufficient approximation the difference of H_c between different specimens at the same temperature could be attributed to differences of transition temperature,* and so in the case of S 7 the critical fields were deduced from those of the other specimens by allowing for its slightly different transition temperature (in this case obtained by extrapolating the h_D, T curve to zero h_D). All the H_c values and the assumed transition temperatures are given in table 1.

For mercury the complication arises that once the mercury has been allowed to melt, it is after the next freezing effectively a different specimen. Longitudinal measurements were taken on H 7, and the critical fields of H 7, H 8 and H 9 appropriate to the transverse experiments were deduced as explained above from estimates of the appropriate transition temperatures. Unfortunately the transition temperatures could not be very accurately determined from the h_D, T curves, and so the H_c values for mercury are not as reliable for mercury as for tin. The possible error in H_c is of course greatest at the highest temperatures, and may be as much as 5 % at 4·05° K and 1 % at 2·1° K. In table 1 we have, in the case of mercury, inserted in brackets alternative values of H_c, and the corresponding assumed transition temperatures, which were found most suitable for fitting theoretical curves (see p. 77).

Comparison of the H_c values for both tin and mercury with those obtained by Andrew from resistance measurements, and other determinations (see II, table 1), shows that our values are on the whole about 1% higher, even at low temperatures (where possible differences in transition temperature are relatively unimportant). The cause of this discrepancy is not clear at present, and requires further investigation.

---

* For S 5, the thinnest specimen, the measured values of H_c were about 1 % higher than those deduced by comparison with S 8 and S 9. This may be evidence of a genuine size dependence of H_c associated with the finite penetration depth.

We are now ready to apply the theory in detail. As explained in III, the theory predicts the transverse magnetization curve of an infinite cylinder of radius r in a complicated implicit form involving H_c and the parameter Δ'/r, where

$$\Delta' = \Delta - \lambda, \tag{1}$$

with

$$\Delta = 8\pi\alpha/H_c^2, \tag{2}$$

where λ is the penetration depth of a magnetic field into the superconducting phase and α is the surface energy per unit area at a boundary between superconducting and normal phases. Although the relation between magnetization and field cannot be expressed explicitly in terms of Δ'/r and H_c, two features of the curve have simple explicit expressions.

For reasons discussed in III (particularly the fact that resistance appears almost simultaneously with the first sharp drop of magnetization) it is likely that only a laminar structure need be considered for the intermediate state, and on this assumption the theory predicts that ρ is given by

$$\rho = \tfrac{1}{2} + 1{\cdot}283\left(\frac{\Delta'}{r}\right)^{\frac{1}{2}} - 1{\cdot}05\left(\frac{\Delta'}{r}\right)^{\frac{3}{4}}. \tag{3}$$

The theory also gives σ explicitly as

$$\sigma = \tfrac{1}{2} + 2{\cdot}2\left(\frac{\Delta'}{r}\right)^{\frac{1}{2}}. \tag{4}$$

There are thus three ways of estimating the value of Δ' from the experimental data, and so constructing the appropriate theoretical magnetization curve.

(i) We may use the experimental values of ρ and compare them with equation (3). In figure 17 we have plotted the experimental values of ρ for tin against Δ'/r, assuming that $\Delta' = 0{\cdot}7 \times 10^{-5}$ cm. at 2·1° K, $1{\cdot}0 \times 10^{-5}$ cm. at 3·00° K and $2{\cdot}0 \times 10^{-5}$ cm. at 3·49° K, and it will be seen that the experimental points lie fairly well on a single curve, which is in fair agreement with the theoretical curve of equation (3), except for high Δ'/r. For mercury the experimental values of ρ are less accurate than for tin and this procedure gives only a rough indication of the value of Δ', suggesting that it is of order $0{\cdot}3 \times 10^{-5}$ cm. at 2·1° K and several times larger at 3·97° K.

(ii) If we plot σ against $r^{-\frac{1}{2}}$, equation (4) predicts a linear graph at any one temperature, with a slope of $2{\cdot}2\Delta'^{\frac{1}{2}}$. This has been done in figure 18 for tin and similar graphs are obtained for mercury; though there is a considerable scatter of the experimental points, Δ' may be estimated roughly, and we find for tin

$$\Delta' = 2{\cdot}0 \times 10^{-5}\text{ cm. at } 2{\cdot}1^\circ\text{ K}, \quad 3{\cdot}6 \times 10^{-5}\text{ cm. at } 3{\cdot}00^\circ\text{ K}$$

and

$$6{\cdot}2 \times 10^{-5}\text{ cm. at } 3{\cdot}49^\circ\text{ K},$$

and for mercury

$$\Delta' = 1{\cdot}2\times10^{-5}\,\text{cm. at } 2{\cdot}1^\circ\,\text{K},\quad 2{\cdot}1\times10^{-5}\,\text{cm. at } 3{\cdot}69^\circ\,\text{K},$$
$$2{\cdot}8\times10^{-5}\,\text{cm. at } 3{\cdot}97^\circ\,\text{K}\quad\text{and}\quad 2{\cdot}1\times10^{-5}\,\text{cm. at } 4{\cdot}05^\circ\,\text{K}.$$

These values are several times larger than those estimated from ρ, and we shall see why this is so directly.

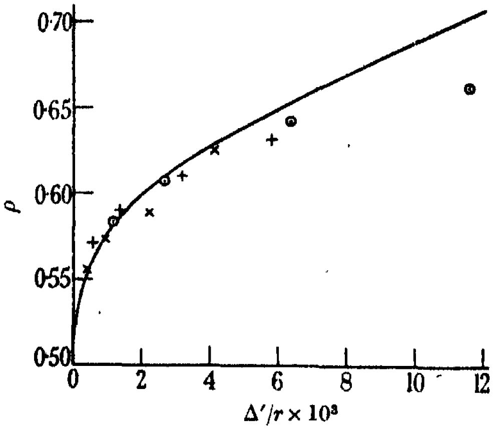


FIGURE 17. Size variation of ρ for tin. Full curve is theoretical (equation (3)). ⊙, 3·5° K assuming $\Delta' = 2\times10^{-5}$ cm.; +, 3·0° K assuming $\Delta' = 10^{-5}$ cm.; ×, 2·1° K assuming $\Delta' = 0{\cdot}7\times10^{-5}$ cm.

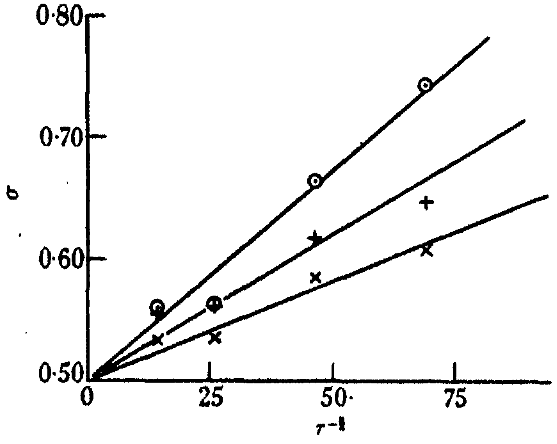


FIGURE 18. Variation of σ with $r^{-\frac{1}{2}}$ for tin. ⊙, 3·5° K; +, 3·0° K; ×, 2·1° K.

(iii) We can attempt by trial and error to choose a value of Δ' at each temperature which will produce theoretical magnetization curves for all the different specimens fitting the experimental data as well as possible. When this attempt was made, a difficulty at once became apparent in the form of a slight inherent discrepancy between theory and experiment. The theory predicts in fact that whatever the value of Δ' (provided it is not too large), the magnetization curve when the whole of the specimen is in the intermediate state should pass through the point $-4\pi I = \frac{2}{3}H_c$,

$h = \frac{3}{4}H_c$, while the experimental curves nearly always pass appreciably above this point. It is possible that this discrepancy arises from simplifying assumptions in the theory, but, whatever its interpretation, it is clear that the value of σ deduced from an experimental curve must always be higher than that corresponding to a theoretical curve in which the magnetization vanishes at about the same field as in the experimental curve. A higher value of σ means, of course, a higher value of Δ' and so it is clear that the estimates of Δ' from σ give an upper limit.

In fitting the curves for tin. the values of H_c used were those obtained from longitudinal measurements, but for mercury it was found that some slight adjustments of the H_c values shown in table 1 greatly improved the fitting. The adjustments amounted essentially to assuming that the transition temperatures were slightly different from those first chosen as explained on p. 74, and affected the high temperature curves most; the adjusted values of T_c and H_c are shown in brackets in table 1 and since they do not differ from those first chosen by more than the likely errors we consider their use quite justified.

The sharp drop in the theoretical curves corresponds to the point where the free energy of the intermediate state becomes equal to that of the superconducting state and, as explained in III, this should really not be a vertical drop but more gradual. Owing to the difficulty of calculating the form of the curve in this region, the drop has been shown vertical in the fitted theoretical curves, but the more gradual fall found experimentally is qualitatively just what would be expected. The fact that the experimental curves usually pass above the point $-4\pi I = \frac{3}{8}H_c$, $h = \frac{3}{4}H_c$ may also be due to the same inadequacy of the theory; if it were possible to calculate the form of the steep transition into the intermediate state properly it might turn out that the transition was not yet quite completed at $h = \frac{3}{4}H_c$ so that the value of $-4\pi I$ should indeed be more than $\frac{3}{8}H_c$. If this inadequacy of the theory is borne in mind, and it is remembered also that the theory probably breaks down in the region of transition from the intermediate to the normal state, it will be seen that the experimental points are in fair qualitative agreement with the theoretical curves.

Apart from the inherent approximations of the theory itself, another possible cause of the lack of detailed agreement may be anisotropy both of the surface energy α and the penetration depth λ, and hence of Δ'. There is indeed experimental evidence that the penetration depth λ is anisotropic (Laurmann & Shoenberg 1947) and there is no reason why α should not be anisotropic too. In this connexion the following rather speculative considerations may be mentioned.

(1) In the case of S 8, which is a single cylinder, there is a suggestion of two separate steep drops in magnetization near the top of each magnetization curve (see, for instance, figure 4). The specimen was examined by cracking off the glass wall and etching with acid, and proved to consist mainly of one large crystal and one smaller one with a different orientation. The experimental curves would thus be explained by supposing that Δ' had different values for the two crystal orientations, and was larger for the larger of the two main crystals. A less marked effect of this kind occurs with H 7 too, but in this case no subsequent investigation could be made of the crystalline character of the specimen.

(2) The fact that the mercury curves are more rounded at the top than the tin ones may be due to the fact that Δ' for mercury is more anisotropic than for tin (though there is no independent evidence for this). If there are many large crystals (assuming each capillary consists of only a few crystals) in a specimen and if they have widely varying values of Δ', this would result in a considerable smoothing out of the theoretical discontinuity for a single value. It is perhaps significant that the fall of magnetization is particularly steep in S 9 which is an extruded tin wire, and so presumably behaves practically isotropically.

The detailed working out of the implications of an anisotropic Δ' is complicated and need not be considered until new experimental data on single crystals is available.

Discussion of the Δ' values

Owing to the inaccuracies of the ρ values given in table 1, and the uncertain interpretation of the σ values, it is likely that the most reliable estimates of Δ' are those obtained by the curve-fitting process. These values are collected in table 2. The fitting of theoretical curves is rather insensitive to the exact choice of Δ' values, but very sensitive to the assumed value of H_c, so the estimates shown in table 2 must be regarded as very rough, especially at high temperatures (for instance, quite a small error in the assumed H_c values would bring Δ' for mercury at 4·05° K below, instead of above, the value at 3·97° K); it is unlikely, however, that our estimates are wrong by more than a factor 2. In table 2 we show also the estimates of Δ' obtained by Andrew from resistance measurements (II). The numerical differences are not greater than can be accounted for by the errors of the two experiments, but Andrew's estimates do not show such a marked temperature variation as ours. Since Andrew's figures are derived essentially from only one feature of the experiment (the reappearance of resistance), while ours are based on the whole magnetization curve, we are inclined to believe that the temperature variation may be really as marked as suggested by our measurements.

Table 2

		Δ' (cm. × 10^5)				estimates of β (cm. × 10^5)	
substance	temp. (° K)	present work	Andrew	penetration depth λ (cm. × 10^5)	$\Delta = \Delta' + \lambda$ (cm. × 10^5)	Désirant & Shoenberg (1948)	Ginsburg (1945)
Sn	3·5	3·7	2·5	1·2	4·9	—	—
	3·0	1·5	1·8	0·7	2·2	—	—
	2·1	0·8	1·7	0·5	1·3	—	—
Hg	4·05	1·1	$<0{\cdot}5$	2·3	3·4	$<0{\cdot}4$	2·4
	3·97	0·9		1·8	2·7		1·9
	3·7	0·7		1·2	1·9		1·0
	2·1	0·4		0·8	1·2		0·7

From our values of Δ' we may estimate Δ, using the values of λ indicated by other experiments; for mercury the values of λ given in table 2 are obtained from experiments of Shoenberg (1940) and Désirant & Shoenberg (1947); for tin only values of

$\lambda(T)-\lambda(2{\cdot}1^\circ\text{K})$ are available (Pippard 1947; Laurmann & Shoenberg 1947), but $\lambda(2{\cdot}1^\circ\text{K})$ can be estimated by extrapolation, using a plausible theoretical form of the temperature variation of λ, and this procedure gives the figures in table 2. Adding Δ' and λ we find the values of Δ given in table 2.

Without a proper theory of superconductivity no interpretation of these values of Δ can be given, but it is of interest to compare them with the values of β, where

$$\beta = 8\pi(\alpha_n-\alpha_s)/H_c^2, \tag{5}$$

and α_n and α_s are the surface free energies per unit area between an insulator and normal and superconducting phases respectively. Only very indirect evidence about β is available from critical field measurements of thin films (Appleyard *et al.* 1939). Ginsburg (1945) and Désirant & Shoenberg (1947, 1948) offer somewhat different interpretations of this evidence, which produce the estimates of β shown in table 2, and it will be seen that though, as might be expected, Δ is of the same order of magnitude as β, it is appreciably greater, so that $\alpha > \alpha_n - \alpha_s$.

We wish to thank Mr E. R. Andrew, Mr A. B. Pippard and Mr H. N. V. Temperley for many valuable discussions and Mr J. M. Look for help with some of the measurements and in the preparation of the diagrams.

REFERENCES

Andrew, E. R. 1947 *Phys. Soc. Camb. Conference Rep.*, vol. II, p. 101.
Appleyard, E. T. S., Bristow, J. R., London, H. & Misener, A. D. 1939 *Proc. Roy. Soc.* A, **172**, 540.
de Haas, W. J., Voogd, J. & Jonker, J. M. 1934 *Physica*, **1**, 281.
Désirant, M. & Shoenberg, D. 1947 *Nature*, **159**, 201.
Désirant, M. & Shoenberg, D. 1948 *Proc. Phys. Soc.* (in the Press).
Ginsburg, V. 1945 *J. Phys. U.S.S.R.* **9**, 305.
Landau, L. 1943 *J. Phys. U.S.S.R.* **7**, 99.
Laurmann, E. & Shoenberg, D. 1947 *Nature*, **160**, 747.
London, F. 1936 *Physica*, **3**, 450.
Misener, A. D. 1938 *Proc. Roy. Soc.* A, **166**, 43.
Peierls, R. 1936 *Proc. Roy. Soc.* A, **155**, 613.
Pippard, A. B. 1947 *Nature*, **159**, 434.
Shoenberg, D. 1937 *Proc. Camb. Phil. Soc.* **23**, 260.
Shoenberg, D. 1938 *Superconductivity*. Cambridge University Press.
Shoenberg, D. 1940 *Proc. Roy. Soc.* A, **175**, 49.
Shoenberg, D. 1947 *Phys. Soc. Camb. Conference Rep.*, vol. II, pp. 85 and 93.

The intermediate state of superconductors ·
II. The resistance of cylindrical superconductors in transverse magnetic fields

By E. R. Andrew, *Stokes Student, Pembroke College and Royal Society Mond Laboratory, University of Cambridge*

(*Communicated by Sir Lawrence Bragg, F.R.S.—Received* 12 *December* 1947
—*Read* 3 *June* 1948)

The resistance of tin and mercury cylinders was measured in transverse magnetic fields between 1·5° K and the superconductivity transition temperature, for cylinder radii from 5×10^{-2} cm. to $1\cdot3\times10^{-3}$ cm. It was found that the field ρH_c, which first caused the appearance of resistance, and which is well known to be greater than $\frac{1}{2}H_c$, increased with decreasing radius approximately in the manner predicted by Landau's theory of the intermediate state. There have been considerable discrepancies between the results of previous workers concerning the temperature dependence of ρ, and on the basis of the present measurements it has been possible to suggest plausible explanations of the discrepancies. It has been shown that the resistance and magnetic induction in the intermediate state are closely correlated as might be expected. Using this correlation in conjunction with the magnetic behaviour predicted on the basis of Landau's theory (see following paper referred to as III), estimates are obtained of the surface energy at the boundary between superconducting and normal phases; it is the finite size of this energy which is responsible for the fact that ρ is greater than $\frac{1}{2}$.

1. Introduction

In part I measurements of the magnetization of cylindrical superconductors in a transverse magnetic field have been described. In this paper, measurements of the resistance of such cylinders will be presented. If a long cylindrical superconductor of circular cross-section is placed in a transverse magnetic field, electrical resistance does not appear when the applied field h is equal to half the critical field H_c, although the field at some part of the surface of the cylinder has then reached the critical value. The applied field, which does eventually cause the appearance of resistance will be referred to as h_1, and the ratio h_1/H_c will be termed ρ_1.

Measurements of ρ_1 by previous workers are summarized below, and show considerable discrepancies:

(*a*) de Haas, Voogd & Jonker (1934), found $\rho_1 = 0\cdot58$ for a monocrystalline tin wire of diameter 0·25 mm. at 2·92° K, using a potentiometer method of measurement.

(*b*) Misener (1938), investigated polycrystalline wires of spectroscopically pure tin, of diameter 0·10 to 0·15 mm. by the same method. He found that ρ_1 was a linear function of temperature, increasing from the value of 0·50 at T_c as the temperature decreased, and having the value 0·57 at 2·92° K in agreement with (*a*). Experiments with indium and lead gave similar results.

(*c*) Mann, Grayson Smith & Wilhelm (1937, 1938) used a persistent current method to study cylinders 3 mm. in diameter. For a single crystal of 'chemically' pure tin, ρ_1 was found to vary with temperature in the manner found by Misener. On the other hand a polycrystalline cylinder of spectroscopically pure tin, diameter 3·65 mm. yielded a value of $\rho_1 = 0{\cdot}58$ independent of temperature.

(*d*) Casimir (1943) has reported numerous measurements on single crystals of diameter about 0·25 mm.* for which ρ_1 had the fairly constant value 0·58, independent of temperature.

In view of these conflicting results, a more thorough examination was undertaken. The investigation has covered a wide range of cylinder diameters in order to test the size-dependence of ρ predicted by Landau (1943), which could not be checked from existing data. Polycrystalline and monocrystalline specimens of tin have been studied, and the investigation has been extended to mercury, both metals being obtainable in a very pure state, and having convenient transition temperatures for use in a liquid helium cryostat. The potentiometer method for resistance measurement was employed, since it gives more information concerning the behaviour of the wire in the intermediate state than the persistent current method, which has the additional disadvantage of being much more sensitive to slight irregularities of shape.

2. THE SPECIMENS

The most important considerations governing the design of the specimens were that they should be as pure and strain-free as possible, and that electrical connexions to them for resistance measurement should involve neither impurity at the joint nor field distortion causing appreciable deviation from the behaviour of the ideal infinite transverse cylinder. The residual resistance, the transition temperature and the width of the temperature transition are a useful guide to purity and freedom from strain.

(*a*) *Tin*

The tin specimens were prepared from two samples, Hilger 10,000 and Johnson & Matthey 12,966 (J. & M.). A spectrographic analysis by the makers showed less than 0·004 % impurity in the former (mainly iron, lead, copper and bismuth), and less than 0·008 % in the latter (mostly lead, with some copper). Hilger 10,000 was the material used by Misener (1938) and for the pure polycrystal of Mann *et al.* (1938).

The tin wires were prepared in three ways: by drawing, by extrusion through a die, and by filling evacuated glass capillaries with molten tin under atmospheric pressure. The last method was most used since it yields specimens with clean surfaces, and can conveniently be applied to very small diameters. The glass could be removed from specimens of large diameter by carefully cracking it off, and, when possible, this was the best method. For narrower diameters the glass could be removed by solution in hydrofluoric acid, but as mentioned later, this method is slightly suspect. Alternatively the glass was left on, except just at the ends, where it was removed to allow leads to be attached.

Single crystals of tin were prepared using a travelling furnace. X-ray oscillation photographs were taken at various points along the length of the axis to confirm their monocrystalline character. The specimens used for this purpose were made in glass capillaries which served to maintain their circular cross-section.

---

* Professor Casimir has kindly informed me of the diameter, the value of which was not given in the paper.

For cylinders of diameter greater than 0·1 mm., potential leads of the same material about 0·1 mm. diameter, were fused with a pinhole gas flame at points 4 to 5 cm. apart. The specimens were examined under a microscope, particularly near the joints, to make sure that their cross-section was uniform, as any irregularity with a demagnetizing coefficient $n > \frac{1}{2}$ would cause the premature appearance of a small amount of resistance. Leads of 44 s.w.g. copper wire were then attached to the four current and potential leads. The assembly was mounted straight on a thin paxolin board using durofix, which was applied to the leads only. If the adhesive was applied to the specimen itself it was found to become strained at low temperatures, causing a change of critical fields. When the glass was left on, there was usually no evidence of strain, the tin probably shrinking away from the glass surface.

For cylinders of diameter less than 0·1 mm., fusion of leads with a pinhole flame was unsatisfactory, while any method of welding or soldering involving another metal would have invited trouble from alloy effects. The cylinder E was therefore laid on paxolin A, and its ends clamped between small brass plates B as indicated in figure 1 (a), current and potential leads being attached to opposite sides of the upper brass plates CC. To reduce the end resistance in the brass a tin foil FF, 20μ thick, was clamped in also, and provided that FF is parallel to the applied field, it causes negligible distortion. The residual end effect was less than 1 % of the normal resistance of the specimen at liquid helium temperatures.

Details of individual specimens are given in table 2. Circularity of cross-section was checked under the microscope, and could be confirmed by comparison of the specimen's behaviour for different directions of the applied transverse field.

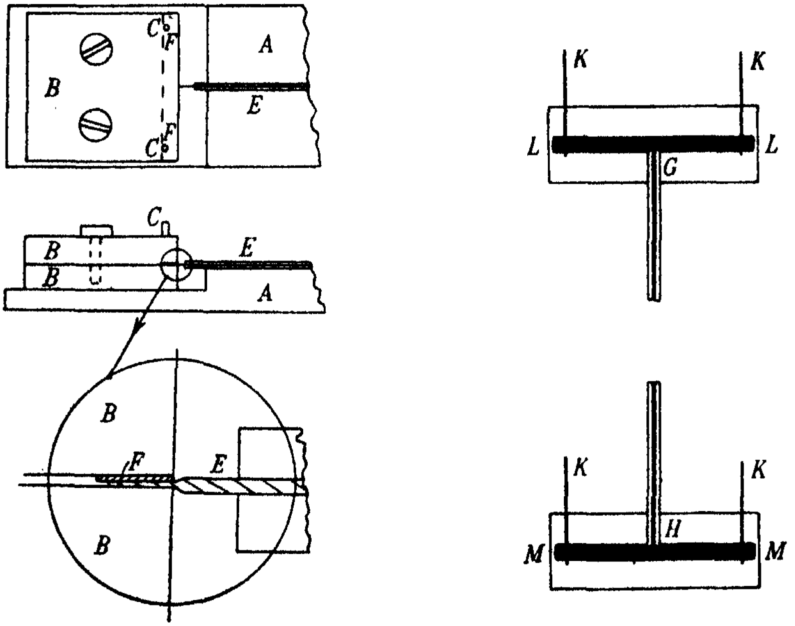
(*b*) *Mercury*

The mercury specimens were prepared from two samples, Hilger 10,673 and 11,023, the spectrographic report of the latter showing 0·0001 % silver as impurity. Here the problem was to devise a method for supplying current and potential leads to a capillary without the use of reservoirs which would cause field distortion. In the arrangement used, figure 1 (b), the glass capillary GH, was sealed into radial holes in two distrene capillaries LL, MM, thus forming two coplanar T-junctions. The seals were made with a solution of distrene in toluene. The distrene capillaries were 14 mm. long and had diameters 1 mm. internal and 5 mm. external. Four platinum wires K, 0·25 mm. diameter, were similarly sealed in, about 1 mm. from each end. Platinum and mercury do not interact appreciably at room temperature. Three of the open ends were sealed, and after filling the system with mercury through the fourth, it too was sealed. This method was used successfully for specimens of diameter down to 25μ. The assembly was mounted on paxolin board and placed vertically in the empty Helium Dewar flask.* By pouring a little liquid air into the outer (liquid air) flask, the mercury cooled slowly from the bottom end and froze without breaking.

---

* In this laboratory the liquid helium is contained in a Dewar vessel surrounded by liquid air in a larger Dewar vessel. The interspace of the inner Dewar flask contains air at low pressure, which acts as a heat conductor at liquid air temperatures. However, at liquid helium temperatures the air solidifies and the vessel becomes 'hard'.

The transverse magnetic field was applied parallel to the reservoirs *LL*, *MM*. Being in a longitudinal field, these remained superconducting until the critical field was reached, so that the four limbs could be used as current and potential leads for the wire *GH*. The additional resistance appearing at the critical field, when the reservoirs ceased to be superconducting, was negligible in comparison with the full resistance of the wire *GH*, which had by then also completely returned. Since the reservoirs were not infinitely long cylinders, they produced a slight distortion of the field at *G* and *H*, but calculation showed that the effect was small enough to be neglected.



(*a*) (*b*)

FIGURE 1. Specimen construction, (*a*) tin for diameters $<0{\cdot}10$ mm.; (*b*) mercury.

3. EXPERIMENTAL METHOD

(*a*) The resistance measurements were made with a potentiometer circuit of the type described by Kapitza & Milner (1937) and Milner (1937). The primary galvanometer was a Cambridge Short Period, whose deflexions were amplified by means of a divided photocell attached to a similar secondary galvanometer. The overall sensitivity was 600 mm./μV for a scale at 1 m. The stability of the spot was such that measurements could be made with an accuracy of 2×10^{-9} V.

(*b*) The transverse magnetic field was produced by a pair of coils arranged in Helmholtz fashion, which gave a transverse field uniform within 1 % over a specimen length of 5 cm. The field along the axis was uniform within $1\frac{1}{2}$ % over the same length, and could be used for measurements in a longitudinal field. For the investigation of the final three specimens *A* 21 to 23, the coil system was slightly modified to produce a greater uniformity of field, namely a transverse field uniform within 0·6 % and a longitudinal field uniform within 0·1 % over a length of 5 cm.

(*c*) The routine examination of the specimens comprised the following:

(i) Resistance measurement at room temperature.

(ii) Temperature transition in absence of applied field.

(iii) Resistance transitions in increasing and decreasing transverse and longitudinal fields at selected temperatures from the transition temperature down to about 1·5° K.

In one case a transverse resistance transition was carried out for a constant value of the magnetic field, the temperature being increased and then decreased again. The resistance-temperature curve was approximately a section of resistance-field curves at constant temperature, and will therefore not be described separately.

The value of the measuring current i A, for transverse transitions, was such as produced a surface field $2i/10r$ of about $H_c/100$. The shift of h_1, with increasing i was found to be approximately in accord with Silsbee's hypothesis, but the field distribution ceased to be symmetrical, which made it desirable to keep the correction small. For each specimen, at one temperature at least, transitions were made for a variety of measuring currents.

(*d*) All temperatures have been calculated on the basis of the 1932 Leiden scale of helium vapour pressures (Keesom 1932).

4. TEMPERATURE TRANSITIONS AND RESISTANCE RATIOS

Tin

The breadth of the temperature transition varied from 0·02° for a polycrystalline specimen to 0·002° for a good single crystal. The transition temperatures, defined as the temperatures at which the resistance had fallen to one-half of its normal value, lay between 3·713 and 3·728° K, with a mean value 3·722° K. The resistance at 4·2° K was almost entirely residual, and its ratio to the room temperature value was small enough to indicate a high degree of purity. For wires of diameter greater than 100μ, the mean ratio was $5{\cdot}5 \pm 0{\cdot}5 \times 10^{-4}$ for Hilger 10,000 and $2{\cdot}2 \pm 0{\cdot}4 \times 10^{-4}$ for J. & M. 12,966. The ratios compare favourably with those found by previous workers, e.g. Sizoo, de Haas & Onnes (1925) and Misener (1938). It will be noticed that the Hilger tin was 2·5 times more resistive than the J. & M. tin, possibly due to the larger iron content. No difference in the superconducting properties however, was noticed. For diameters less than 100μ increasing residual resistance was found, which probably has its origin in the mean free path of the conduction electrons becoming comparable with the wire diameter (see for instance Lovell 1938).

Mercury

The breadth of the temperature transition was always less than 0·02° K and transition temperatures lay between 4·160 and 4·180° K with a mean value 4·170° K. There was no apparent difference in the low temperature resistivities of the two samples used. As the Debye temperature is very low for mercury, the ideal resistance still predominated at 4·2° K, the resistance being a mean fraction $4{\cdot}5 \times 10^{-4}$ of the room temperature (liquid) value. At 1·5° K, the resistance in a transverse field H_c, had fallen to 4×10^{-5} of the room temperature value for diameters greater than 100μ. The residual resistance would be still less than this as the magneto-resistance effect is quite important here. For diameters less than 100μ the resistance showed higher values as with tin.

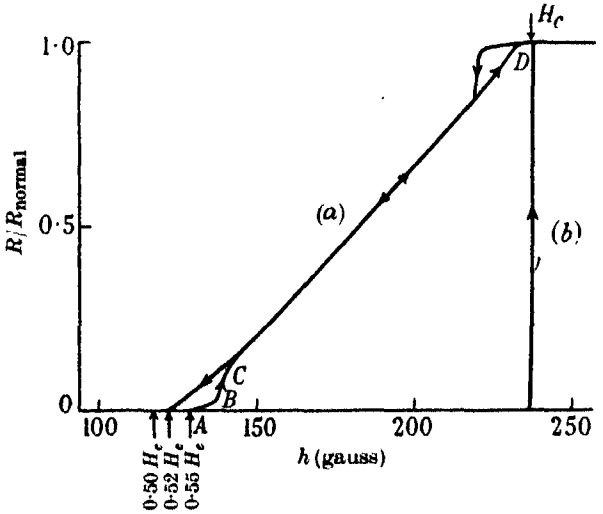


FIGURE 2. Typical tin resistance transitions. (*a*) In increasing and in decreasing transverse fields; (*b*) in increasing longitudinal field. Diameter 0·40 mm. at 1·666° K; measuring current 200 mA.

5. GENERAL FEATURES OF RESISTANCE TRANSITIONS

(*a*) *Tin*

A typical resistance transition is shown in figure 2, and serves to indicate the general features. The arrows indicate the transitions in increasing and decreasing fields. Resistance first appears at a quite definite value h_1, and grows slowly at first, *AB*. It then rises more sharply *BC*, finally increasing almost linearly with field (provided, as will be seen later, that the measuring current is not too small), until just before the critical field is reached, where there is a slight rounding at *D*, the whole normal resistance having returned just before $h = H_c$ (H_c determined from longitudinal measurements, discussed later).

With any one specimen there is of course a spread of critical fields corresponding to the finite temperature transition width, amounting to two or three gauss in the worst cases. This spread contributes to the rounding *D*, and is most in evidence at the highest temperature, setting the limit of approach of useful measurements to the transition temperature. In addition there is a genuine rounding, roughly proportional to H_c and therefore not caused by spread of critical fields, which increases with diminishing wire diameter, as indicated in figure 3. It is apparent therefore that the extrapolation method for determining the critical field from transverse transitions suggested by Misener (1938), can only be valid for large diameters. The rounding at D corresponds to the final tail of the magnetization curves (see I, e.g. figure 2) and the extrapolated field value corresponds to the size-dependent h_D mentioned in I. The theoretical interpretation of these, and other features will be discussed in III.

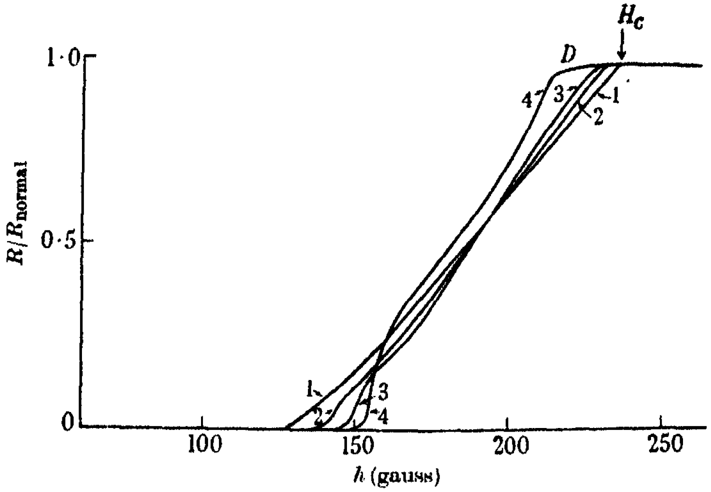


FIGURE 3. Size dependence of resistance transition for tin at 1·66° K. Cylinder diameters: (1) 1·05 mm., (2) 0·15 mm., (3) 0·062 mm., (4) 0·027 mm.; with measuring currents 400, 50, 15 and 6 mA respectively.

The resistance-field curves reflect the main characteristics of the magnetization curves of tin cylinders, described in I. The exact correlation of the two was confirmed by a combined experiment on specimen A 23 for which magnetization measurements were carried out by Désirant & Shoenberg (who refer to the specimen as S 9) simultaneously with the resistance measurements. The specimen consisted of a 27 cm. length of extruded tin wire of diameter 0·15 mm., wound bifilarly on a cylindrical distrene former 0·84 cm. diameter, pitch 2 × 1 mm., comprising ten turns and a diameter. These dimensions represent a compromise between angle of rise of the helix (4°) and adequate spacing of the turns (mean demagnetization coefficient modified from $\frac{1}{2}$ to 0·493), compatible with the space available in the magnetization apparatus.

A typical magnetization curve for this specimen is shown in I, figure 1. The corresponding B-h curves have been deduced from the magnetization curves, and are shown with the resistance curves in figure 4, the curves in increasing and in decreasing fields being separated for clarity. It is not to be expected that the resistance curves will follow the B-h curves exactly, since different arrangements of the phases in the intermediate state could produce different resistance values for the same B. It will be seen, however, from figure 4 that there was a close correspondence between the pairs of curves. Penetration of induction commenced at K, indicated by the appreciable departure from linearity of the magnetization curve. Resistance did not, however, appear until l (one part in 10^4 readily detectable). In the interval KL there is penetration without resistance. The rapid rise of resistance mn corresponds to the sharp increase in induction MN, indicated by the sharp drop of magnetization. The linear portions NO, no correspond, and the rounding of the resistance transition op, has its counterpart in the B-h curve, indicated by the tail of the magnetization curve. This correspondence was of considerable value in confirming the interpretation of the resistance curves, and the main features of both magnetization and resistance curves find explanation in III. In these simultaneous experiments magnetization measurements were made both with and without the resistance measuring current flowing, but no significant effect of this current was found.

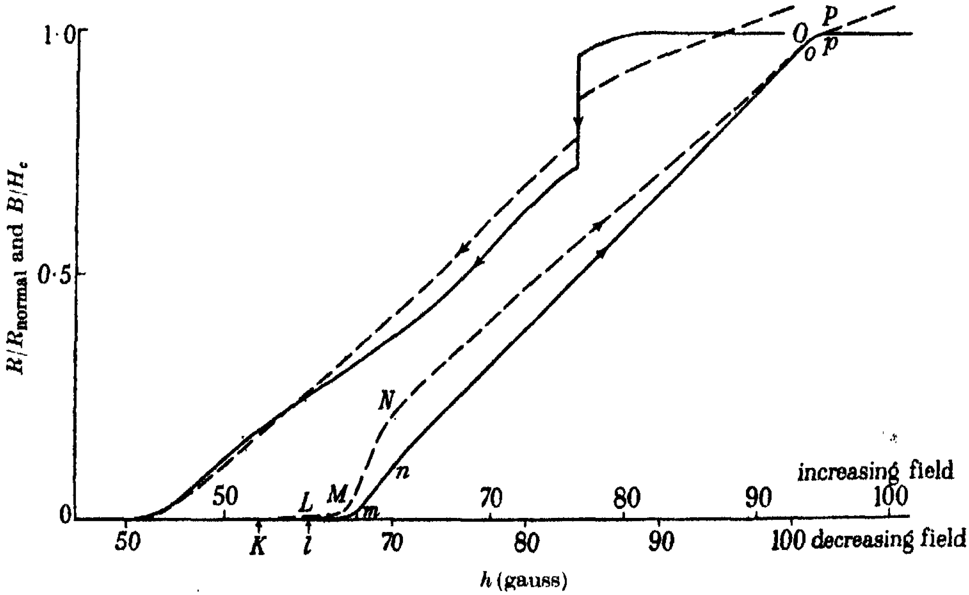


FIGURE 4. Resistance and magnetic induction curves for the tin specimen A 23 at 2·998° K, measuring current 30 mA. Full line, resistance; broken line, magnetic induction.

It is the ratio $\rho_1 = h_1/H_c$ determined by first appearance of resistance (A figure 2, l figure 4) which is to be compared with the results of previous workers, but it is the ratio $\rho_2 = h_2/H_c$ determined by the sharp fall of magnetization and sharp rise of resistance (B figure 2, Mm figure 4) which corresponds to the transition of the whole cylinder into the intermediate state, as will be seen in III, and which is to be compared with the theory.

The critical fields H_c used in computing these ratios, were determined from longitudinal transitions on the same specimens, and mean values are given in table 1, where they are compared with other published values. Interpolated values of the mean critical fields used by Désirant & Shoenberg in I are also shown. These interpolated values from I are about 1 % higher than those obtained in the present experiments, which agree quite well with the other published values. The comparison is only fair, however, if the mean transition temperatures are the same. The mean values of T_c from I (table 1), are in fact only 0·002° greater than those found in the present experiments for both tin and mercury, but it should be pointed out that in I, T_c is obtained by an extrapolation of the H_c-T curve, which for mercury is stated to be a very approximate procedure. Moreover, there is evidence that the extrapolation method gives lower values of T_c than the resistance temperature transition indicates. It is therefore likely that there is no real discrepancy here and that the differences with tin (for which the estimates of H_c in I are more accurate), are due to a small mean displacement of the threshold field curve, possibly caused by the fact that in I all the specimens were in glass capillaries which may produce small strains, whereas in the present experiments only half the specimens were in capillaries. There is in any case a possible error of ½ % in the field calibrations in both series of experiments. Discrepancies due to this cause, and to real differences in H_c values, will not affect the values of ρ.

TABLE 1. CRITICAL FIELD VALUES H_c FROM LONGITUDINAL RESISTANCE TRANSITIONS

tin

temperature (°K)	H_c (gauss)	de Haas & Engelkes (1937)	Daunt & Mendelssohn (1937)	Désirant & Shoenberg I
3·48	34	34	34	35
3·20	72	70	72	—
2·96	103	103	100	104
2·53	154	155	150	—
2·15	196	195	191	198
1·66	237	238	235	—

mercury

temperature (°K)	H_c (gauss)	Daunt & Mendelssohn (1937)	Misener (1940)	Désirant & Shoenberg I
4·00	32	34	32	34
3·50	122	121	122	123
3·00	200	200	201	—
2·50	268	269	268	—
2·00	322	320	322	325
1·50	362	359	363	—

First then we examine values of ρ_1, which are given in table 2 for a number of cylinders of diameter ranging from 0·105 down to 0·0027 cm. and for temperatures from 3·5 to 1·5° K. It is apparent that there is no great temperature variation of ρ_1, in contrast with the results of Misener, but that for a given specimen, ρ_1 is almost constant, in accord with the Leiden experiments (Casimir 1943) and the Canadian experiments for spectroscopically pure tin (Mann *et al.* 1938). The mean value of ρ_1, for each specimen is seen to increase from 0·54 for a diameter of 0·105 cm. to 0·65 for a diameter of 0·0027 cm. Closer examination shows a slight *decrease* in ρ_1, with temperature of the order of 2 % from the highest to the lowest temperatures. In § 7 the results of other workers are compared with the present measurements, and explanations are offered of those which do not agree.

TABLE 2. VALUES OF ρ_1 AND ρ_2 FOR TIN

ρ_1 in ordinary type, ρ_2 in italics

specimen	details (see notes below)	diameter (mm.)	$\frac{R_{4\cdot2}}{R_{\mathrm{room}}}\times 10^4$	3·48° K	3·20° K	2·96° K	2·53° K	2·15° K	1·67° K	mean†
A 14	J, M, drawn	1·05	2·7	*	$0{\cdot}54_6$	$0{\cdot}54_3$	$0{\cdot}55_1$	$0{\cdot}53_2$	$0{\cdot}54_6$	$0{\cdot}54_4$
				$0{\cdot}54_9$	*$0{\cdot}55_1$*	*$0{\cdot}54_3$*	*$0{\cdot}55_6$*	*$0{\cdot}54_6$*	*$0{\cdot}54_8$*	*$0{\cdot}54_9$*
A 6	J, P, capillary; all glass cracked off	0·47	2·1	$0{\cdot}57_7$	$0{\cdot}55_2$	$0{\cdot}56_3$	$0{\cdot}55_4$	$0{\cdot}54_9$	$0{\cdot}54_8$	$0{\cdot}55_7$
				$0{\cdot}58_6$	*$0{\cdot}56_0$*	*$0{\cdot}56_6$*	*$0{\cdot}56_6$*	*$0{\cdot}57_0$*	*$0{\cdot}56_3$*	*$0{\cdot}56_5$*
A 7	J, P, capillary; ends cracked off	0·40	1·9	*	$0{\cdot}57_5$	$0{\cdot}57_5$	$0{\cdot}55_4$	$0{\cdot}54_7$	$0{\cdot}55_2$	$0{\cdot}56_1$
				$0{\cdot}63_1$	*$0{\cdot}59_2$*	*$0{\cdot}58_3$*	*$0{\cdot}57_2$*	*$0{\cdot}58_2$*	*$0{\cdot}57_9$*	*$0{\cdot}58_1$*
A 13	J, M, capillary; ends cracked off	0·17	2·1	$0{\cdot}57_4$	$0{\cdot}53_0$	$0{\cdot}55_9$	$0{\cdot}56_1$	$0{\cdot}55_8$	$0{\cdot}56_1$	$0{\cdot}55_7$
				$0{\cdot}60_8$	*$0{\cdot}59_7$*	*$0{\cdot}59_5$*	*$0{\cdot}59_7$*	*$0{\cdot}58_5$*	*$0{\cdot}59_3$*	*$0{\cdot}59_3$*
A 4	H, P, extruded	0·15	6·0	$0{\cdot}55_7$	—	$0{\cdot}58_1$	$0{\cdot}58_3$	$0{\cdot}56_8$	$0{\cdot}56_1$	$0{\cdot}57_6$
				$0{\cdot}61_7$	—	*$0{\cdot}61_0$*	*$0{\cdot}61_2$*	*$0{\cdot}61_0$*	*$0{\cdot}59_6$*	*$0{\cdot}60_7$*
A 23	H, P, extruded	0·15	5·1	$0{\cdot}57_8$	—	$0{\cdot}59_0$	—	$0{\cdot}57_8$	—	$0{\cdot}58_2$
				$0{\cdot}62_0$	—	*$0{\cdot}61_5$*	—	*$0{\cdot}59_5$*	—	*$0{\cdot}60_5$*
A 9	J, P, capillary; ends cracked off	0·13	2·5	$0{\cdot}59_3$	$0{\cdot}58_6$	$0{\cdot}58_9$	$0{\cdot}57_7$	$0{\cdot}57_3$	$0{\cdot}57_9$	$0{\cdot}58_1$
				$0{\cdot}63_7$	*$0{\cdot}62_1$*	*$0{\cdot}61_7$*	*$0{\cdot}61_5$*	*$0{\cdot}61_3$*	*$0{\cdot}60_5$*	*$0{\cdot}61_4$*
A 15	J, 2S, capillary; glass dissolved off	0·12	2·5	*	$0{\cdot}57_5$	$0{\cdot}58_2$	$0{\cdot}57_4$	$0{\cdot}56_7$	$0{\cdot}57_1$	$0{\cdot}57_4$
				$0{\cdot}63_7$	*$0{\cdot}61_6$*	*$0{\cdot}61_2$*	*$0{\cdot}61_3$*	*$0{\cdot}62_3$*	*$0{\cdot}61_7$*	*$0{\cdot}61_6$*
A 21	J, P, capillary; ends cracked off	0·062	6·0	$0{\cdot}61_4$	$0{\cdot}61_9$	$0{\cdot}61_1$	$0{\cdot}60_5$	$0{\cdot}61_2$	$0{\cdot}59_3$	$0{\cdot}60_9$
				$0{\cdot}64_8$	*$0{\cdot}63_6$*	*$0{\cdot}64_5$*	*$0{\cdot}62_8$*	*$0{\cdot}63_7$*	*$0{\cdot}62_5$*	*$0{\cdot}63_4$*
A 22	J, P, capillary; ends cracked off	0·027	9·1	$0{\cdot}64_6$	$0{\cdot}65_0$	$0{\cdot}66_1$	*	$0{\cdot}64_5$	$0{\cdot}63_0$	$0{\cdot}64_6$
				$0{\cdot}67_3$	*$0{\cdot}66_3$*	*$0{\cdot}67_8$*	*$0{\cdot}67_6$*	*$0{\cdot}66_7$*	*$0{\cdot}66_7$*	*$0{\cdot}67_0$*

J, H: specimen prepared from Johnson & Matthey 12,966, Hilger 10,000 respectively. M: monocrystalline. P: polycrystalline. 2S: two single crystals.

* Owing to a discontinuous rise ρ_1 could not be estimated.

† As explained in the text, only the five lowest temperatures are used in computing the mean of ρ_2.

We now turn to values of ρ_2. It is possible to estimate h_2 to about 1 %. Values of ρ_2 thus obtained are also given in table 2. Here again, as with ρ_1, there is little variation with temperature below 3·25° K, while at the highest temperature 3·48° K, ρ_2 is on the average about 3 % higher. At this temperature, however, the accuracy of the measurements is least, as the inhomogeneity of the critical field is always an appreciable fraction of the critical field (34 gauss).

There is less temperature variation of ρ_2 than found for $\rho = h_A/H_c$ in I, so that for the five lower temperatures it is justifiable to take a mean value for each specimen for the purpose of comparison with the size-variation predicted by the theory. This mean is also given in table 2. Again it is evident that ρ_2 increases continuously as the cylinder diameter decreases. In III formulae for ρ_2 have been derived on the basis of models suggested by Landau (1943). For the thread-like model the relation is

$$\rho_T = \tfrac{1}{2} + 2{\cdot}31\left(\frac{\Delta'}{r}\right)^{\frac{1}{2}} - 1{\cdot}38\left(\frac{\Delta'}{r}\right)^{\frac{3}{4}}, \tag{1}$$

while for the laminar model it is

$$\rho_L = \tfrac{1}{2} + 1{\cdot}283\left(\frac{\Delta'}{r}\right)^{\frac{1}{3}} - 1{\cdot}05\left(\frac{\Delta'}{r}\right)^{\frac{2}{3}}, \tag{2}$$

where $$\Delta' = \Delta - \lambda \quad \text{and} \quad \Delta = \alpha\frac{8\pi}{H_c^2}.$$

λ is the penetration depth of a magnetic field into the superconducting phase, α is the energy per unit area of surface between superconducting and normal phases and r is the cylinder radius.

In figure 5 the mean values of ρ_2 given in table 2 for tin are plotted against the cylinder radius, the latter on a logarithmic scale, together with curves calculated from equations (1) and (2) assuming suitable values of Δ'. The experimental values fit equation (2) better than (1), suggesting that the intermediate state commences with a laminar form, and thus supporting the mechanism of penetration suggested in III. Fair agreement with the theoretical curve is obtained assuming $\Delta' = 1{\cdot}7 \times 10^{-5}$ cm. The greatest deviation occurs for the cylinder of smallest radius ($r = 80\Delta'$), which suggests that the approximations of the theory, which require $r \ll \Delta'$, are beginning to fail here.

The small temperature variation of ρ_2 in the range covered by the five lowest temperatures, mentioned above, corresponds to a decrease of Δ' from $2{\cdot}0 \times 10^{-5}$ cm. at 3·20° K to $1{\cdot}6 \times 10^{-5}$ cm. at 1·67° K. The larger values of ρ_2 at 3·48° K correspond to $\Delta' = 2{\cdot}5 \times 10^{-5}$ cm., but as already mentioned the accuracy in the determination of ρ_2 is much less here, and deviations from (2) are greater. As mentioned in I, the values of ρ_2 agree approximately with those obtained from the magnetization curves, and the differences are probably due to a combination of the following causes: latitude in the determination of both h_A and h_2; the discrepancies in H_c already mentioned; differences inherent in the specimens themselves, such as small strains.

The specimens so far mentioned, and others which were investigated less systematically, showed no significant difference in behaviour between monocrystalline and polycrystalline cylinders, except in so far as the former had sharper temperature transitions, and consequently a smaller critical field range. However, it will be seen from (1) and (2) that Δ' is a rapidly varying function of ρ, and the scatter of the experimental ρ values could easily mask an anisotropy in Δ' of the kind suggested in I.

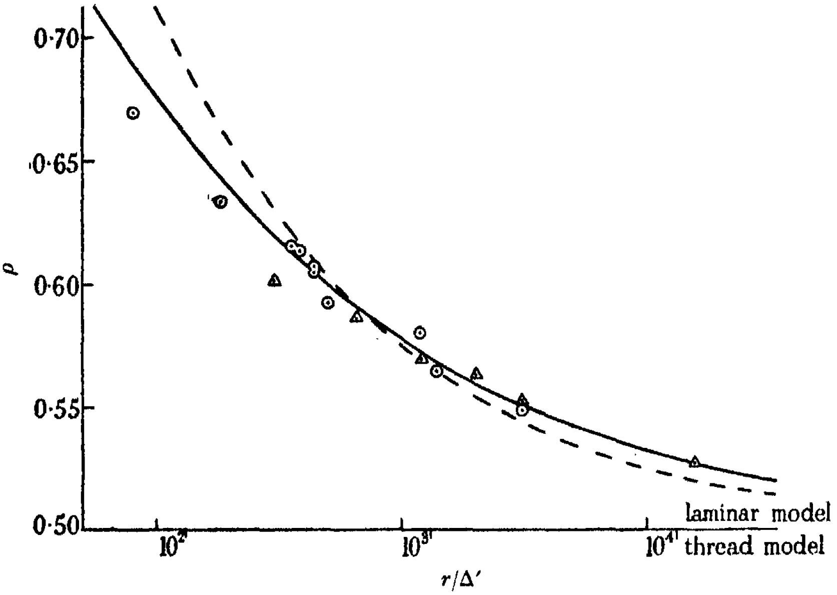


FIGURE 5. Experimental ρ values plotted with the theoretical curves of laminar model (full line) and thread model (broken line). ⊙ Tin with $\Delta' = 1{\cdot}7 \times 10^{-5}$ cm. for laminar model and $2{\cdot}2 \times 10^{-5}$ cm. for thread model. △ Mercury with $\Delta' = 0{\cdot}5 \times 10^{-5}$ cm. for laminar model and $0{\cdot}65 \times 10^{-5}$ cm. for thread model.

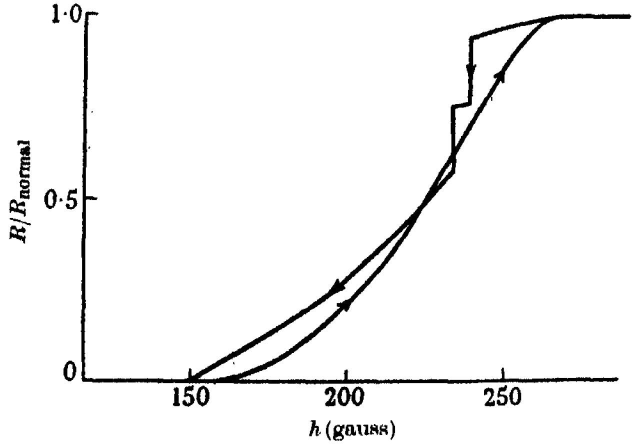


FIGURE 6. Typical resistance curve for a mercury cylinder. Diameter 0·065 mm., 2·495° K.

(*b*) *Mercury*

A typical resistance curve is shown in figure 6, which illustrates the difficulty of quantitative deduction for this metal. Penetration appears to take place more gradually than in tin, as evidenced also by the rounded tops of the magnetization curves in I, which unlike those for tin show no very sharp drop. Similarly the resistance shows no sharp rise but increases more smoothly than with tin. It is therefore impossible to derive a value h_2 from the transitions. Values of h_1, the field which produces the first trace of resistance, are used to obtain the values of ρ_1, given in table 3. As with tin, there is little variation of ρ_1, with temperature, while the mean value for each specimen increases with decreasing diameter. The experimental values fit relation (2) quite well, figure 5, with a value of $\Delta' = 5{\cdot}0 \times 10^{-6}$ cm. which must be regarded as a lower limit if as with tin $\rho_1 < \rho_2$. The values of ρ obtained from the magnetization curves are stated in I to be only of sufficient accuracy to give a rough indication of the value of Δ', so that there is no significant difference in the indications of the two methods.

TABLE 3. VALUES OF ρ_1 FOR MERCURY

specimen	material	diameter (mm.)	$\frac{R_{1\cdot5}}{R_{\text{room}}} \times 10^5$ $(h = H_c)$	4·00° K	3·50° K	3·00° K	2·50° K	2·00° K	1·50° K	mean
B6	H. 11,023	0·79	3·3	$0{\cdot}54_0$	$0{\cdot}53_8$	$0{\cdot}52_5$	$0{\cdot}53_9$	$0{\cdot}53_7$	$0{\cdot}54_8$	$0{\cdot}53_7$
B3	H. 10,673	0·31	5·0	$0{\cdot}54_2$	$0{\cdot}54_4$	$0{\cdot}54_8$	$0{\cdot}56_3$	$0{\cdot}55_6$	$0{\cdot}56_9$	$0{\cdot}55_4$
B1	H. 10,673	0·20	3·4	$0{\cdot}57_4$	$0{\cdot}54_8$	$0{\cdot}55_6$	$0{\cdot}56_9$	$0{\cdot}56_5$	$0{\cdot}56_9$	$0{\cdot}56_4$
B2	H. 10,673	0·12	4·9	$0{\cdot}58_2$	$0{\cdot}58_4$	$0{\cdot}56_8$	$0{\cdot}55_7$	$0{\cdot}56_3$	$0{\cdot}56_7$	$0{\cdot}57_0$
B4	H. 11,023	0·065	7·2	$0{\cdot}58_1$	$0{\cdot}58_6$	$0{\cdot}58_3$	$0{\cdot}59_6$	$0{\cdot}58_4$	$0{\cdot}59_2$	$0{\cdot}58_7$
B5	H. 11,023	0·030	11·5	$0{\cdot}59_7$	$0{\cdot}60_9$	$0{\cdot}60_6$	$0{\cdot}60_8$	$0{\cdot}59_8$	$0{\cdot}59_5$	$0{\cdot}60_2$

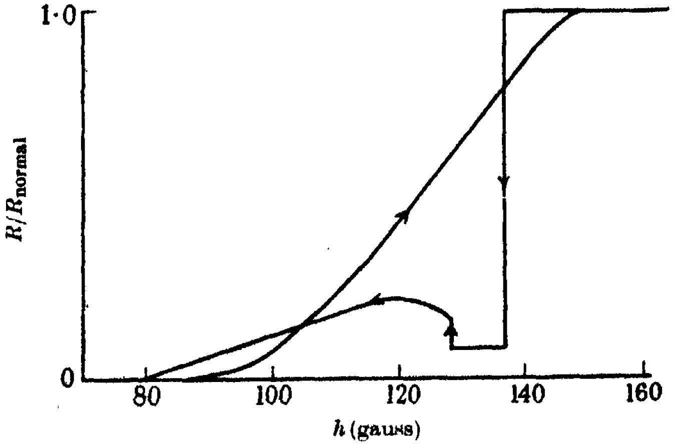


FIGURE 7. Hysteresis cycle for tin specimen A 15, two single crystals, 0·12 mm. diameter, 2·535° K, measuring current 10 mA.

6. HYSTERESIS EFFECTS

(a) *Resistance-field effects*

The main features are illustrated by the cycles in figures 2, 4, 6 and 7, and have their counterparts in the magnetization curves in I. Possible explanations of these effects are given in III. There was always supercooling on the return curve, sometimes with a steep continuous fall (e.g. figure 2), but usually with a discontinuous drop to the intermediate state at a field value which was not perfectly reproducible. The magnitude of the supercooling varied widely from one specimen to another but in general was greater for specimens of smaller diameter. The return curve often coincided with that for increasing field, but was sometimes rather higher and occasionally as in figure 7, dropped well below, on one case to zero, an effect which has been noticed previously by de Haas & Voogd (1931). This effect need not imply a return to the superconducting state, but merely an arrangement of the intermediate state with a superconducting path between the ends, as would be the case if laminae were set parallel to the wire axis. The return curves did not show the same degree of concavity at the lower resistance end as did curves in increasing field, resistance disappearing at lower values of h, which will be termed h_3. We may define another ratio $\rho_3 = h_3/H_c$, which like ρ_1 and ρ_2 has little temperature dependence. The size variation of ρ_3 is, however, much smaller, and ρ_3 was usually between 0·52 and 0·55 for all sizes. There was frequent evidence that the metal was not always in its most stable condition, as the following examples indicate. Discontinuous transitions, such as that from the normal state to the intermediate state were occasionally initiated by making and breaking the measuring current in balancing the potentiometer. Another curious hysteresis effect which was sometimes encountered is shown in figure 7. In the combined resistance and magnetization experiment, mentioned earlier, the magnetic field was on one occasion switched off when the metal was in the intermediate state with B about $\frac{1}{2}H_c$. When the field was suddenly switched on again, the resistance was found to have a different (smaller) value than before, although the magnetization was unchanged. If, however, the field was switched off and increased gradually to its former value as in a normal cycle, the original value of resistance was again obtained.

(*b*) *Current effects*

As found by de Haas *et al.* (1934) and by Misener (1939) and illustrated in figures 8 (*a*) and 9 (*a*) for mercury and tin respectively, the curves in increasing field were more concave for weaker measuring currents, as if in the limit of zero measuring current, the transverse and longitudinal transitions would become identical. For thinner specimens a limiting curve was found, smaller measuring currents producing no further concavity. This limiting curve became less concave with decrease of diameter, and for 30μ had become practically linear. In fact at that diameter there was scarcely any current effect at all. Similar effects were found in a decreasing field, as instanced in figure 8 (*b*). On a laminar picture, one may suppose that laminae lie almost parallel to the axis for large diameters and weak measuring currents (or electric fields), and turn normal to the axis for larger electric fields; for small diameters, however, the laminae are almost perpendicular to the axis even for weak electric fields.

(*c*) *Current hysteresis*

A new hysteresis effect was observed when the current was varied for constant temperature and magnetic field. Suppose a transition in a transverse field has been partially completed, and the state of the cylinder is given by P, figure 9 (*a*). Then if the magnetic field is held constant, and the measuring current increased, the resistance increases, as given by Q, R, S. When a large current has been reached, such that the transition curve reached is linear, there is little further increase, figure 9 (*b*). If now the measuring current is decreased, the resistance remains constant at the value determined by the largest current, instead of decreasing again

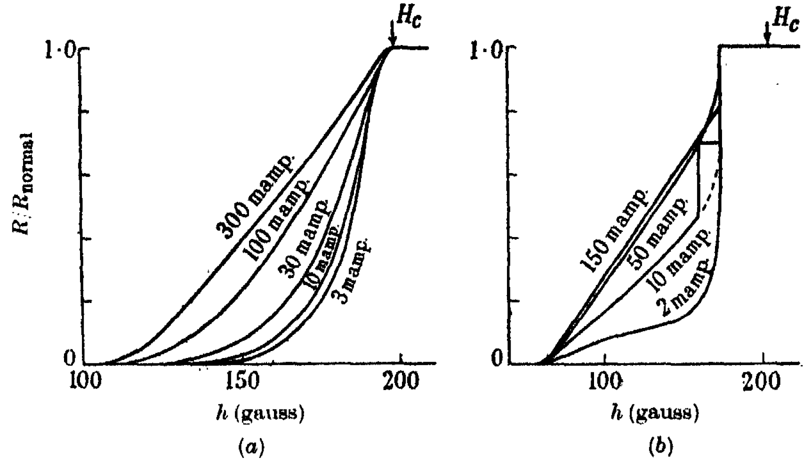


Figure 8. Current effects. (*a*) For transition in an increasing field, mercury, diameter 0·31 mm., 3·007° K; (*b*) for transition in a decreasing field, tin, diameter 0·15 mm. 2·530° K.

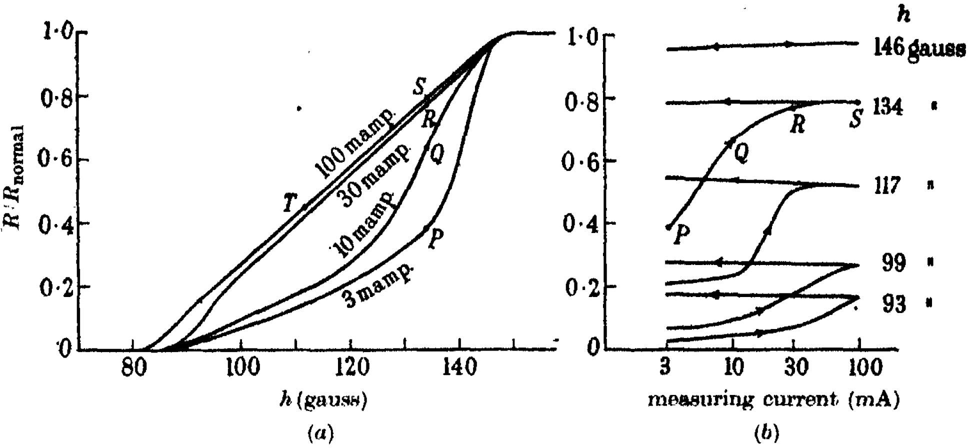


Figure 9. Current hysteresis. (*a*) Resistance transition for constant measuring current; (*b*) resistance transitions for constant transverse field. Monocrystalline tin specimen, A 13, diameter 0·17 mm., 2·530° K.

to R, Q, P. Similar hysteresis effects are shown in figure 9 (*b*) for other values of the constant transverse field. In another experiment, the resistance transition with a current of 100 mA was partially completed as indicated by T, figure 9 (*a*). The current was reduced to 10 mA, resistance remaining unchanged. The transition was then completed using 10 mA and was found to follow the 100 mA curve. Hence it seems that once an arrangement of phases in the intermediate state has been impressed by a higher current, subsequent reductions of current do not modify it.

7. COMPARISON WITH PREVIOUS WORK

These experiments have shown that the ratio ρ_1 for spectroscopically pure tin showed little temperature variation, in agreement with the results reported by Mann *et al.* (1938), and by Casimir (1943). Such small temperature variation as was found was in the opposite sense to the strong temperature variation found by Misener (1938). The experiments have also shown that ρ_1 is a function of the cylinder diameter, having the value 0·57 for a diameter of 0·25 mm. (interpolated from table 2), in satisfactory agreement with the value 0·58 given by de Haas *et al.* (1934) and by Casimir (1943). On the other hand, for a cylinder 3·65 mm. in diameter, an extrapolation of the experimental values in the light of the theoretical relation (2), predicts $\rho_1 = 0·525$ which is well below the value of 0·58 found by Mann *et al.* (1938). The following possible explanations may be offered of the two discordant experiments just mentioned.

(*a*) *The results of Misener* (1938)

Curious results obtained with one specimen A 16, suggested an explanation of Misener's results. This polycrystalline drawn tin wire of diameter 1·05 mm. of the J. & M. 12,966 sample, was compared with the wire A 14 of table 2, from which it differed apparently only in being a single crystal. Its resistance behaved, however, very differently from that of A 14; the main features of difference were:

(i) h_1 differed considerably from the values of A 14, see table 4, the fractional discrepancy increasing at lower temperatures.

(ii) The transition took the form of figure 10, resistance increasing rapidly at first to about 0·8 of its normal value, and then increasing slowly until full resistance had returned, which did not occur until a field about twice as high as H_c at the lowest temperature.

(iii) At 1·66° K, a longitudinal transition was also measured and found to be very broad, over a range 370 to 610 gauss ($H_c = 236$ gauss).

TABLE 4. THE ANOMALOUS SPECIMEN A16 COMPARED WITH A14

temperature (° K)	h_1 for A 16 (gauss)	h_1 for A 14 (gauss)	H'_c for A 16 (gauss)	H_c for A 16 gauss)	$\rho' = h_1/H'_c$
3·48	20	—	36	33	0·55$_8$
3·20	41·5	38·5	72	70	0·58
2·96	64	56	103	103	0·62
2·52	112	84	158	152	0·71
2·15	157	104	198	195	0·79
1·66	222	128	250	235	0·89

If apparent values of critical field, H'_c, are obtained by Misener's extrapolation method indicated in figure 10, they are found to be not very different from the critical field values of A 14 (both given in table 4). Moreover, if apparent values of ρ are obtained from $\rho' = h_1/H'_c$, they are found to increase almost linearly with $\Delta T = T_c - T$ (table 4 and figure 11). The slope is greater than Misener's, and similar to that obtained with 'chemically' pure tin by Mann & Wilhelm (1937).

The broad transitions implied impurity or strain, but the resistance at room-temperature and at 4·2° K had normal values, with $R_{4\cdot2}/R_{\text{room}} = 2\cdot4 \times 10^{-4}$. A spectrographic analysis of a portion of the wire did not show measurably greater impurity than the material from which the wire had been prepared. These facts, coupled with the shape of the transition curve, suggested that the bulk of the wire was pure, behaving normally, but that there was an impure surface layer. This layer would not affect the resistance appreciably above the transition temperature, but would have a higher critical field than the pure substance in the usual alloy manner. The alloy layer would shunt the pure core, especially at fields greater than the true H_c, when the core is in the normal state. h_1 would be determined by the impure layer, and H_c' by the pure core, and hence ρ_1' would increase with temperature if the critical field curve of the impure layer rose more linearly than that of the pure metal. To explain the shape of the resistance transition, the effective depth of the layer would have to be a few per cent of the diameter. This hypothesis was con-

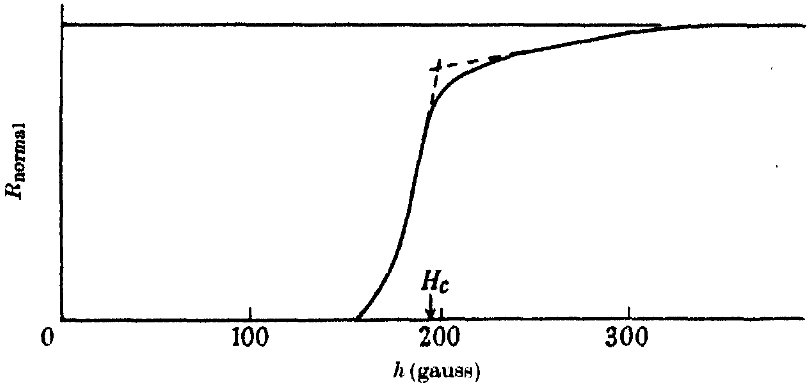


FIGURE 10. Transverse transition for the anomalous specimen A 16, at 2·147° K.

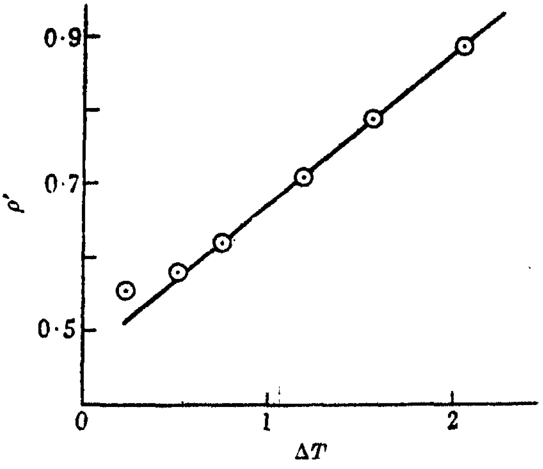


FIGURE 11. Apparent temperature variation of ρ for the anomalous specimen A 16.

firmed by dissolving away the surface of the wire in hydrochloric acid, until its diameter was 10 % smaller, when it was found to behave quite normally, giving a sharp longitudinal transition at the correct field, a temperature independent value of ρ, and a full return of resistance at H_c in a transverse field.

Hence it seems possible that the wire used by Misener also had an impure surface layer, the impurity being smaller than for the above specimen A 16, the history of which, unlike that of the other specimens, was not quite certain, as it had not been drawn in this laboratory. The fact that surface purity is very important was also shown by the fact that tin wires prepared in glass capillaries subsequently dissolved off with hydrofluoric acid, sometimes acquired an anomalous surface layer. In this case the main effect was to broaden the temperature transition considerably. Dissolving away the surface in hydrochloric acid removed this anomaly also.

(*b*) *The results of Mann et al.* (1938)

In the persistent current method, the intermediate state is approached under entirely different conditions than those prevailing in the resistance measurement method. The magnetic field is highly asymmetrical, so that magnetization cannot be uniform. The authors show that the magnetic induction penetrates the cylinder on the side of high field at $h = 0{\cdot}44H_c$. This penetration could cause a reduction of the demagnetization coefficient, such that the persistent current i did not become zero until $0{\cdot}58H_c$. A simple calculation on this basis accounts satisfactorily for the i-h curve found. Nevertheless penetration at fields $h < 0{\cdot}50H_c$ should not be stable, and it is possible that the time effects found had not completely died away even after 60 sec. as the authors assumed. Whether this explanation be true or not, the fact that the value 0·58 is the same as that of ρ_1 found in the original resistance measurements of de Haas *et al.* (1934), appears to be coincidental.

Mann *et al.* also showed that the addition of small quantities of impurity considerably affected the ratio ρ_1, and it is therefore likely that the curious results they obtained with the less pure tin described as 'chemically' pure, are to be attributed to the presence of bulk impurities.

I acknowledge with pleasure the advice received from Dr Shoenberg, who suggested this problem. My thanks are also due to Dr C. L. Smith who prepared the tin single crystals, and to Mr E. Laurmann who carried out the spectographic analysis mentioned in the text. Most of the work was performed whilst holding an award from the Department of Scientific and Industrial Research.

REFERENCES

Appleyard, E. T. S., Bristow, J. R., London, H. & Misener, A. D. 1939 *Proc. Roy. Soc.* A, **172**, 540.
Casimir, H. B. G. 1943 *Physica*, **10**, 757.
Daunt, J. G. & Mendelssohn, K. 1937 *Proc. Roy. Soc.* A, **160**, 127.
Désirant, M. C. & Shoenberg, D. 1947 *Nature*, **159**, 201.
Ginsburg, V. 1945 *J. Phys. U.S.S.R.* **9**, 305.
de Haas, W. J. & Engelkes, A. D. 1937 *Physica*, **4**, 325.
de Haas, W. J. & Voogd, J. 1931 *Commun. Phys. Lab. Univ. Leiden*, no. 212*c*.

de Haas, W. J., Voogd, J. & Jonker, J. M. 1934 *Physica*, **1**, 281.
Kapitza, P. & Milner, C. J. 1937 *J. Sci. Instrum.* **14**, 165.
Keesom, W. H. 1932 *Commun. Phys. Lab. Univ. Leiden*, suppl. 71*d*.
Landau, L. 1943 *J. Phys. U.S.S.R.* **7**, 99.
Lovell, A. C. B. 1938 *Proc. Roy. Soc.* A, **166**, 270.
Mann, K. C. & Wilhelm, J. O. 1937 *Trans. Roy. Soc. Canada* (iii), **31**, 19.
Mann, K. C., Grayson Smith, H. & Wilhelm, J. O. 1938 *Phys. Rev.* **54**, 758.
Milner, C. J. 1937 *Proc. Roy. Soc.* A, **160**, 207.
Misener, A. D. 1938 *Proc. Roy. Soc.* A, **166**, 43.
Misener, A. D. 1939 *Proc. Camb. Phil. Soc.* **35**, 95.
Misener, A. D. 1940 *Proc. Roy. Soc.* A, **174**, 262.
Sizoo, C. J., de Haas, W. J. & Onnes, H. K. 1925 *Commun. Phys. Lab. Univ. Leiden*, no. 180*c*.

The intermediate state of superconductors
III. Theory of behaviour of superconducting cylinders in transverse magnetic fields

By E. R. Andrew, *Stokes Student, Pembroke College and Royal Society Mond Laboratory, University of Cambridge*

(*Communicated by Sir Lawrence Bragg, F.R.S.—Received* 12 *December* 1947 —*Read* 3 *June* 1948)

The magnetic behaviour of cylindrical superconductors in transverse magnetic fields is developed for Landau's branching laminar model of the intermediate state, and for a branching thread model. The main features of the magnetization and resistance curves presented in parts I and II receive qualitative explanation, and expressions are derived for quantitative comparison of experiment with theory.

1. Introduction

In connexion with the experimental work reported in parts I and II of this series, the behaviour of models of the intermediate state put forward by Landau (1943), has been developed. Landau described a laminar model of the intermediate state, mentioning that when the magnetic induction B is small, a structure consisting of threads of normal metal would probably be more favourable thermodynamically than that of laminae. Such a model was not actually described in the paper, and since the results for both models will be required, a thread structure is first of all discussed here. The magnetic behaviour of both models is then developed, and used to construct a possible explanation for the observed behaviour of cylindrical superconductors in a transverse magnetic field.

2. Thread model

Consider a plate of thickness L in the intermediate state, situated in a magnetic field h normal to its surfaces, figure 1. In the centre of the plate the magnetic induction is carried by identical cylindrical threads of normal metal parallel to the

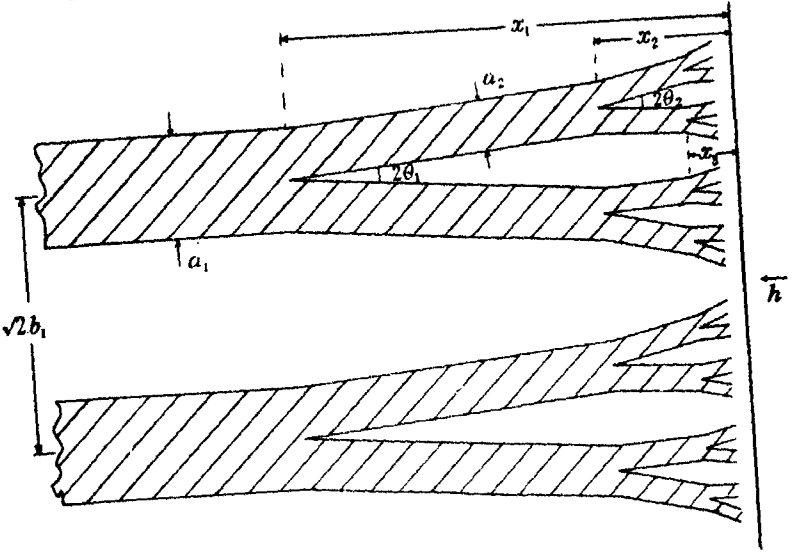


FIGURE 1. Diagonal section of thread system.

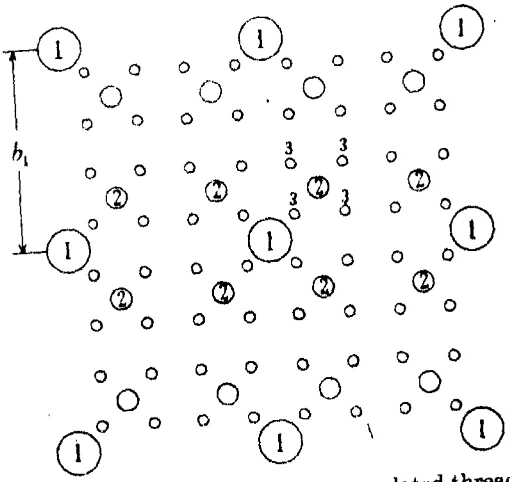


FIGURE 2. Surface pattern of extrapolated threads.

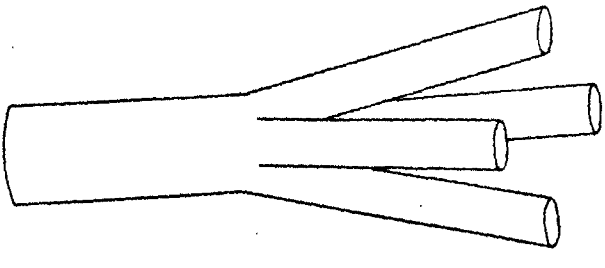


FIGURE 3. Typical branching process.

applied field, and embedded in superconducting metal, the induction in each being H_c. The threads are taken as circular in cross-section, and arranged in square array such that if produced to meet the surfaces, they would intersect as the large circles 1, figure 2. As in the laminar model, to overcome difficulties of stable entry of the induction into the metal, without infinite fields at corners, the threads are supposed to divide successively each time into four equal threads, two in one plane containing h, and two in a perpendicular plane containing h. A typical branching process is shown in figure 3. If after the first branching process the threads were produced to meet the surface, an array of circles 2, figure 2, would be obtained, conditions being adjusted so that this array is a similar square one, but with half spacing. After a second branching, the circles 3 are obtained, and so on, the array being reproduced half-size after each subdivision. A section through the plate in a plane containing the axis of one of the principal threads and a diagonal of figure 2, gives figure 1. After many subdivisions a 'mixed phase' is reached at the surface, having an almost uniform density of magnetic induction.

The detailed calculations, which are exactly parallel to those of Landau for the laminar case, are outlined in the appendix. They are accomplished in the following five steps:

(i) The diameters a_n of the threads after $n-1$ branching processes are related by the condition of constant flux through a tube of induction.

(ii) Minimization of free energy for a single branching process, yielding a relation between the diameter a_n of the thread, the semi-angle of branching θ_n and the surface energy between the two phases $\alpha = \frac{H_c^2}{8\pi}\Delta$ per unit area. The quantity Δ thus defined has the dimensions of a length.

(iii) The reproduction of the arrays of figure 2 half-size at each sub-division imposes a relation on the distances x_n of the branching processes from the surface.

(iv) The mean value of B for the metal as a whole determines the spacing b_1 of the main threads relative to their diameter a_1.

(v) Finally the total free energy of the system is minimized with respect to the remaining variable a_1.

The results of these steps are given in table 1, together with those of the laminar model for comparison. The following points are noted:

(*a*) The free energy f before the minimization of step (v), is a slowly varying function of a_1 (7), with a shallow minimum such that a factor of two in a_1 merely increases the free energy by 10 % of the second (small) term of (8). This is relevant when later we consider how exactly the intermediate state is likely to conform to the structure of minimum free energy.

(*b*) The effect of penetration into the superconducting phase has been neglected. For those superconducting regions which are much thicker than the penetration depth λ, the effect is to reduce the radius of each normal thread, and the half-thickness of each normal lamina, by λ. A corresponding decrease in free energy $\frac{H_c^2}{8\pi}\lambda$ per unit interphase surface area is produced, so that wherever Δ has been used in these calculations it should be replaced by $\Delta' = \Delta - \lambda$, a result given by Ginsburg

(1945). This simple correction cannot, however, apply close to the surface where the superconducting regions have subdivided to a thickness of the order of λ, and where the mixed state may be considered to begin. Using the relations of table 1, it may be shown, however, that the depth of the mixed state thus defined, is of the order of $\lambda^{\frac{3}{2}}/\Delta^{\frac{1}{2}}$ for both models. Since $\lambda \sim \Delta \sim 10^{-5}$ cm., the depth of the mixed phase is small enough to justify the use of Ginsburg's correction for all cases in which these models

TABLE 1

thread model	laminar model (Landau 1943)	
$a_n = \dfrac{a_1}{2^{n-1}}$	$a_n = \dfrac{a_1}{2^{n-1}}$	(1)
$\theta_n = 2\sqrt{\left(\dfrac{\Delta}{a_n}\right)}$	$\theta_n = \sqrt{\left(\dfrac{2\Delta}{a_n}\right)}$	(2)
$\theta_n = 2^{\frac{1}{2}(n-1)}\theta_1$	$\theta_n = 2^{\frac{1}{2}(n-1)}\theta_1$	(3)
$x_{n-1} = 2\sqrt{2}\,x_n$	$x_{n-1} = 2\sqrt{2}\,x_n$	(4)
$2x_1\theta_1 = \dfrac{b_1}{\sqrt{2}} - \dfrac{a_1}{2}$	$2x_1\theta_1 = \dfrac{b_1}{2}$	(5)
$\dfrac{a_1}{b_1} = \sqrt{\left(\dfrac{4B}{\pi H_c}\right)}$	$\dfrac{a_1}{a_1+b_1} = \dfrac{B}{H_c}$	(6)
$f = \dfrac{H_cBL}{4\pi} + \dfrac{H_cB\Delta}{4\pi} \times \left[\dfrac{2L}{a_1} + \dfrac{a_1^{\frac{1}{2}}}{(\sqrt{2}-1)}\sqrt{\left(\dfrac{\pi H_c}{\Delta B}\right)}(1-\beta)\right]$	$f = \dfrac{H_cBL}{4\pi} + \dfrac{H_cB\Delta}{4\pi} \times \left[\dfrac{L}{a_1} + \sqrt{\dfrac{a_1}{\Delta}}\,\dfrac{1}{(\sqrt{2}-1)}\,\dfrac{H_c-B}{B}\right]$	(7)
$\Phi = \dfrac{4\pi f_{\text{min.}}}{L} = H_cB + \dfrac{3\pi^{\frac{1}{3}}B^{\frac{2}{3}}H_c^{\frac{4}{3}}(1-\beta)^{\frac{2}{3}}}{2^{\frac{1}{3}}[(\sqrt{2}-1)]^{\frac{2}{3}}}\left(\dfrac{\Delta}{L}\right)^{\frac{2}{3}}$ $= H_cB+\phi \qquad \beta = \sqrt{\left(\dfrac{2B}{\pi H_c}\right)}$	$\Phi = \dfrac{4\pi f_{\text{min.}}}{L} = H_cB + \dfrac{3B^{\frac{1}{3}}H_c(H_c-B)^{\frac{2}{3}}}{2^{\frac{2}{3}}[(\sqrt{2}-1)]^{\frac{2}{3}}}\left(\dfrac{\Delta}{L}\right)^{\frac{2}{3}}$ $= H_cB+\phi$	(8)
$a_{1\,\text{min.}} = 2^{\frac{4}{3}}[\sqrt{(2-1)}]^{\frac{2}{3}}\left(\dfrac{B\Delta}{\pi H_c L}\right)^{\frac{1}{3}}(1-\beta)^{-\frac{2}{3}}L$	$a_{1\,\text{min.}} = 2^{\frac{2}{3}}[\sqrt{(2-1)}]^{\frac{2}{3}}\left(\dfrac{\Delta}{L}\right)^{\frac{1}{3}}\left(\dfrac{B}{H_c-B}\right)^{\frac{2}{3}}L$	(9)
$x_{1\,\text{min.}} = \dfrac{\sqrt{(2-1)}}{2\sqrt{2}}L = 0{\cdot}146L$	$x_{1\,\text{min.}} = \dfrac{(\sqrt{2}-1)}{2\sqrt{2}}L = 0{\cdot}146L$	(10)
$\dfrac{d\phi}{dB} = \dfrac{2^{\frac{2}{3}}\pi^{\frac{1}{3}}B^{-\frac{1}{3}}H_c^{\frac{4}{3}}(1-\frac{3}{2}\beta)}{[(\sqrt{2}-1)]^{\frac{2}{3}}(1-\beta)^{\frac{1}{3}}}\left(\dfrac{\Delta}{L}\right)^{\frac{2}{3}}$	$\dfrac{d\phi}{dB} = \dfrac{H_c(H_c-3B)}{2^{\frac{2}{3}}[(\sqrt{2}-1)]^{\frac{2}{3}}B^{\frac{2}{3}}(H_c-B)^{\frac{1}{3}}}\left(\dfrac{\Delta}{L}\right)^{\frac{2}{3}}$	(11)

Where L = thickness of the plate.
H_c = critical field.
B = mean magnetic induction.
$\Delta = 8\pi\alpha/H_c^2$, where α is the surface energy per unit interphase area.
θ_n = semiangle of branching at the nth junction.
x_n = distance of the nth junction from the plate surface.
a_n = diameter of the normal threads, or the thickness of the normal laminae, before the nth branching process.
b_1 = distance between axes of the main threads (figure 2), or the thickness of the superconducting laminae.
f = free energy of the system per unit area of plate before the minimum of step (v).
$\Phi/4\pi = f_{\text{min.}}/L$ = minimum free energy per unit volume.
$\phi = \Phi - H_cB$, so that $H = \partial\Phi/\partial B = H_c + d\phi/dB$.

are applicable. Consideration of the stability of superconductors in the normal state for fields greater than H_c leads to the conclusion (London 1935), that $\Delta > \lambda$, so that $\Delta' = \Delta - \lambda$ is a positive quantity. This simple correction assumes that the penetration depth is a function of temperature only. If, however, λ is also a function of magnetic field as Ginsburg suggests may be the case, the formulation of the problem would be considerably modified.

(*c*) The assumption of $B = H_c$ in the normal regions of the intermediate state has been made, but strictly we should put $B = H$, and determine H thermodynamically. This would make the mathematics much more complicated, but because $H = H_c + d\phi/dB$, where $d\phi/dB$ is a first order correcting term, in general small compared with H_c, the modification to the theory is of the second order, and its omission does not cause serious error.

3. THE POINT OF ENTRY INTO THE INTERMEDIATE STATE

Landau (1943) has shown that the free energy of an infinite cylinder of circular cross-section in the superconducting state situated in a transverse magnetic field h, including the field energy, is equal to that for the cylinder in the intermediate state for a value of $B = B_t$ given by:

$$\phi - B_t\left(\frac{d\phi}{dB}\right)_t = \tfrac{1}{2}B_t^2. \tag{12}$$

The corresponding value of applied field $h = h_t$ being given by:

$$2h_t = H_c + B_t + \left(\frac{d\phi}{dB}\right)_t. \tag{13}$$

When ϕ and $d\phi/dB$ are substituted in (12) from table 1, equations in B_t are obtained, which in conjunction with (13) yield a solution for $\rho = h_t/H_c$ in series form, as shown in the appendix, the important terms being:

Thread model:

$$\rho_T = \tfrac{1}{2} + \left(\frac{2\sqrt{\pi}}{(\sqrt{2}-1)}\right)^{\frac{1}{2}}\left(\frac{\Delta'}{L}\right)^{\frac{1}{2}} - \frac{1}{\sqrt{(2\pi)}}\left(\frac{2\sqrt{\pi}}{(\sqrt{2}-1)}\right)^{\frac{3}{4}}\left(\frac{\Delta'}{L}\right)^{\frac{3}{4}}$$

$$= \tfrac{1}{2} + 2{\cdot}926\left(\frac{\Delta'}{L}\right)^{\frac{1}{2}} - 1{\cdot}966\left(\frac{\Delta'}{L}\right)^{\frac{3}{4}}. \tag{14}$$

Laminar model:

$$\rho_L = \tfrac{1}{2} + \frac{5}{8}\left(\frac{4}{(\sqrt{2}-1)}\right)^{\frac{2}{5}}\left(\frac{\Delta'}{L}\right)^{\frac{2}{5}} - \frac{1}{4}\left(\frac{4}{(\sqrt{2}-1)}\right)^{\frac{4}{5}}\left(\frac{\Delta'}{L}\right)^{\frac{4}{5}}$$

$$= \tfrac{1}{2} + 1{\cdot}548\left(\frac{\Delta'}{L}\right)^{\frac{2}{5}} - 1{\cdot}534\left(\frac{\Delta'}{L}\right)^{\frac{4}{5}}. \tag{15}$$

The question now arises: what is the value of L? The thickness of the cylinder increases from zero at the edges to $2r$, the diameter, in the centre. If we require the value of h at which the total free energies are equal, then the appropriate mean value of L must be taken. If on the other hand we require the value of h for which the free energies appertaining to any part of the cylinder first become equal, then the diametrical value $L = 2r$ must be taken. The significance of this choice will be discussed later, but for the present putting $L = 2r$ in (14) we find:

$$\rho_T = \tfrac{1}{2} + 2{\cdot}07\left(\frac{\Delta'}{r}\right)^{\frac{1}{2}} - 1{\cdot}19\left(\frac{\Delta'}{r}\right)^{\frac{3}{4}}. \tag{16}$$

Equation (16) is of the same form as that given by Landau (1943), (disregarding the final term which in practice is small). However, in (16) the coefficient of $(\Delta'/r)^{\frac{1}{2}}$ is about five times larger in Landau's equation. A corresponding calculation for a sphere gives $\rho_T = \frac{2}{3} + 2{\cdot}32(\Delta'/r)^{\frac{1}{2}}$, and here again the coefficient is about five times larger than that given by Landau. Since Landau does not give a detailed calculation it will be assumed that our coefficient is correct and that Landau has made a numerical error. An alternative possibility is that entirely different models have been used. In this connexion it should be stated that two other thread models have been studied, one similar to this, but having threads of square cross-section, and the other having threads of circular cross-section but branching into three instead of into four at each junction, triangular arrays being reproduced $1/\sqrt{3}$ smaller at each subdivision, just as in figure 2 square arrays are reproduced half-size. These two models gave free energy values very little different from those of the model described here and yield relations for ρ_T of the form of equation (16), the coefficients of $(\Delta'/r)^{\frac{1}{2}}$ being 2·20 and 2·12 respectively.

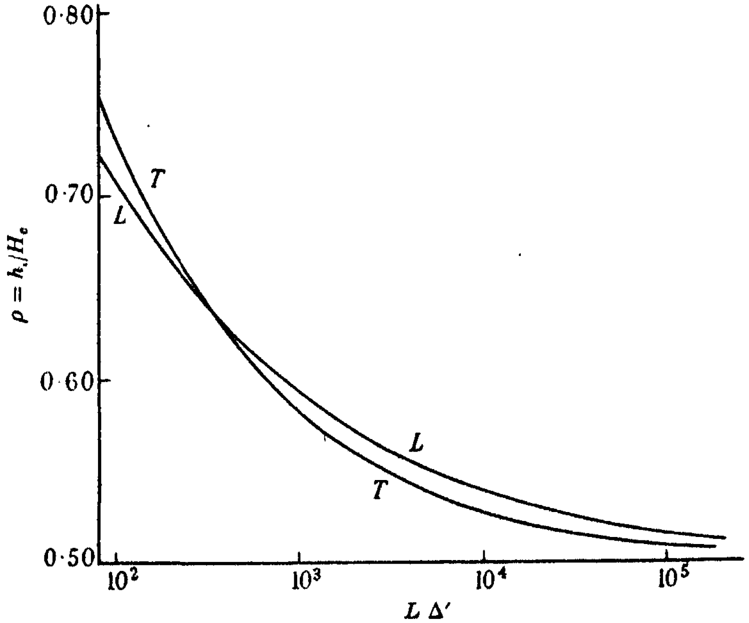


FIGURE 4. The variation of ρ with L/Δ'; curve T for thread model, curve L for laminar model.

ρ_T and ρ_L are plotted as a function of L/Δ' in figure 4, and it will be seen that $\rho_L > \rho_T$ for $L > 300\Delta'$. In view of the approximations made it is clear that these calculations can only be reliable for $L \gg \Delta'$. However, as will be seen later these structures will probably only be encountered for values of which yield $\rho < 1/\sqrt{2}$. From equations (14) and (15) it follows that for $\rho < 1/\sqrt{2}$, Δ'/L is less than about 10^{-2}, and the corresponding angles θ_1 are thus less than about 25°, and small enough for the calculations to be a reasonable approximation.

4. MAGNETIC BEHAVIOUR

Commencing with the three general relations:

$$H = h - 4\pi n I, \quad B = H + 4\pi I, \quad H = \frac{\partial \Phi}{\partial B}, \tag{17}$$

where the demagnetizing coefficient n is $\frac{1}{2}$ in this case, the complete magnetic behaviour of the cylinder may be described. The B-h relation for the intermediate state takes the form $B = 2h - H_c - d\phi/dB$, so that for the thread model:

$$\frac{B}{H_c} = 2\frac{h}{H_c} - 1 - M\left(\frac{B}{H_c}\right)^{-\frac{1}{3}} \frac{(1 - \frac{3}{2}\beta)}{(1-\beta)^{\frac{1}{3}}} \tag{18}$$

and for the laminar model:

$$\frac{B}{H_c} = 2\frac{h}{H_c} - 1 - N\left(\frac{B}{H_c}\right)^{-\frac{2}{3}} \frac{(1 - 3B/H_c)}{(1 - B/H_c)^{\frac{1}{3}}}, \tag{19}$$

where $M = \left[\frac{2\sqrt{\pi}}{(\sqrt{2}-1)}\frac{\Delta'}{L}\right]^{\frac{2}{3}}$ and $N = \left[\frac{1}{2[(\sqrt{2}-1)]}\frac{\Delta'}{L}\right]^{\frac{2}{3}}$.

Now the free energy of a magnetized ellipsoid, together with the field energy, reckoned per unit volume of the ellipsoid, F, is given by (Landau 1943):

$$4\pi F = \Phi - \frac{HB + hB - hH}{2}.$$

Hence for the cylinder in the normal state

$$\frac{8\pi F_n}{H_c^2} = 1,$$

while in the superconducting state

$$\frac{8\pi F_s}{H_c^2} = 2\frac{h^2}{H_c^2},$$

and in the intermediate state

$$\frac{8\pi F_i}{H_c^2} = \frac{2\phi}{H_c^2} + \frac{B}{H_c} + \frac{h}{H_c}\left(1 - \frac{B}{H_c}\right) + \frac{(h-B)}{H_c^2}\frac{d\phi}{dB}.$$

In figure 5, $8\pi F/H_c^2$ is plotted for the normal and superconducting states, and for the intermediate states with $\Delta'/L = 1{\cdot}5 \times 10^{-3}$ (both models). The free energy curves for the thread and laminar models are not very different, and in the limit of $\Delta'/L \to 0$, all models must have the same free energy:

$$\frac{8\pi F_{i,0}}{H^2} = \frac{4h}{H} - \frac{2h^2}{H^2} - 1.$$

This is also shown in figure 5, and it will be noticed that the curve touches that of the superconducting state at A, where $h = \frac{1}{2}H_c$, and that of the normal state at B, where $h = H_c$. For any non-zero value of Δ'/L, however, the intermediate state curves intersect that of the superconducting state at C and D with finite angles. If phase transitions take place at C and D, which correspond to $h/H_c = \rho_T$ and ρ_L respectively, the finite angles of intersection imply an accompanying discontinuous change of magnetization. For $\Delta'/L = 0$, however, transitions at A and B will not be accompanied by a change of magnetization, since the angles of intersection are in this case zero.

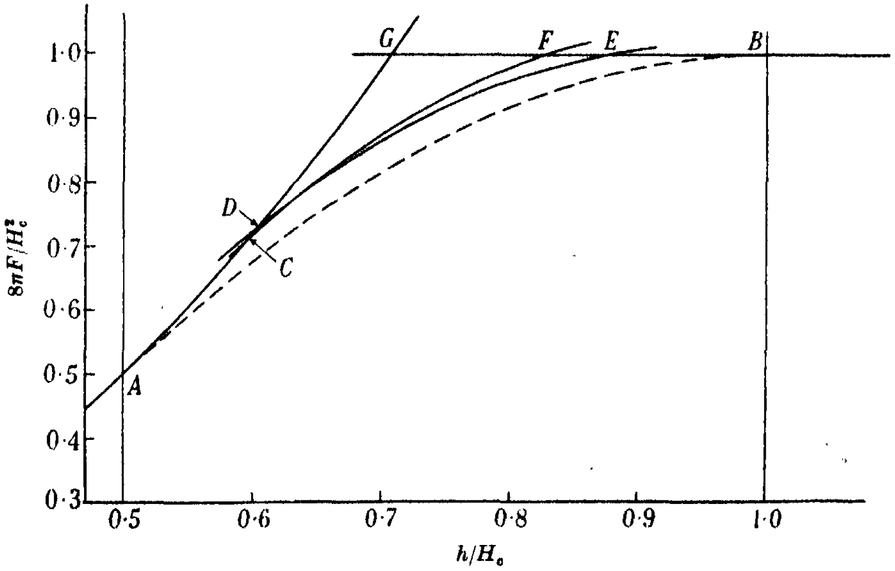


FIGURE 5. Free-energy curves for a transverse cylinder.

AG Superconducting state.
BG Normal state.
DE Intermediate state, laminar model, } $\Delta'/L = 1{\cdot}5 \times 10^{-3}$.
CF Intermediate state, thread model, }
AB Intermediate state, any model, $\Delta'/L = 0$ (broken curve).

Transitions from one state to another do not necessarily take place when the second state becomes more favourable thermodynamically, but only if a mechanism is available for the process in the course of which the free energy does not increase. Failing such a mechanism, the first state continues in a metastable condition, as instanced by the supercooling of a liquid. If, however, possible mechanisms of transition do exist whenever free energy curves meet, then the B-h and I-h curves for the transverse cylinder follow directly from equations (17), (18) and (19), the transitions points being obtained from figure 5. The complete curves are shown in figure 6 where only the laminar form of intermediate state is illustrated for simplicity. Also the additional facts that $B = 0$ and $I = -h/2\pi$ in the superconducting state, and that $B = h$ and $I = 0$ in the normal state are used. Since the point of intersection E, figure 5, occurs for h appreciably less than H_c, the transition into the normal state should take place here despite the applied field being less than the critical field. However, the possibility of another form of intermediate state taking over in the interval EB is not excluded. The transitions at C and E would not be quite discontinuous as shown in figure 6 since L varies from 0 to $2r$, but would nevertheless be fairly sharp since half the volume is contained within the region defined by $L/2r > 0{\cdot}92$.

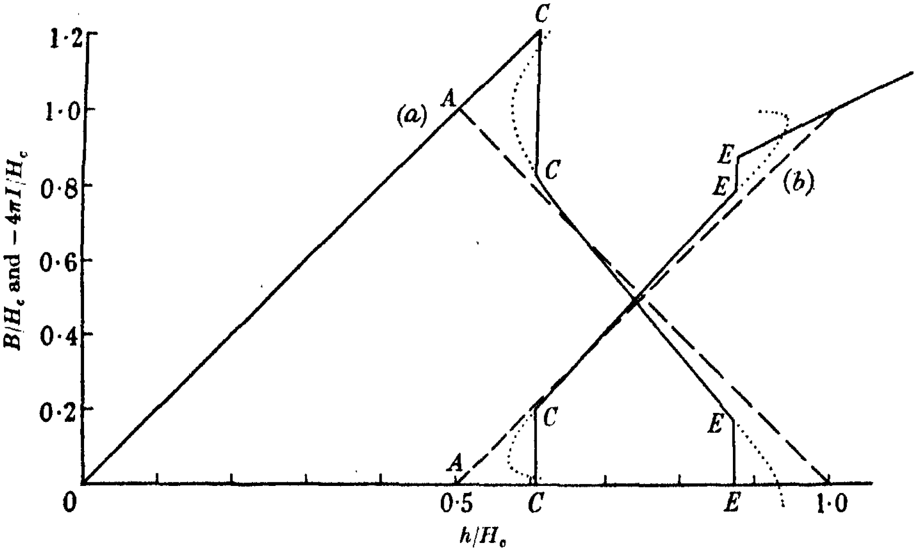


FIGURE 6. Thermodynamically most favourable B-h and I-h curves for a transverse cylinder with $\Delta'/L = 1{\cdot}5 \times 10^{-3}$. Broken lines for $\Delta'/L = 0$. (a) Magnetization. (b) Magnetic induction.

In figure 6 the B-h and I-h curves are continued beyond the transition points as dotted lines, and are seen to show turning points, with I apparently becoming positive. These curious features are associated with the approximation, previously mentioned, of taking H as H_c in the normal regions, since the difference term $d\phi/dB$ becomes large as B tends to the limiting values 0 and H_c, as indicated by (11) table 1.

It should be noticed from G, figure 5, that the superconducting state can never be stable for $h > H_c/\sqrt{2}$, quite irrespective of considerations of the intermediate state. If no free energy curve of any form of intermediate state intersects that of the superconducting state before $H_c/\sqrt{2}$ then a transition directly into the normal state should occur. In fact at this value of h, the points C, E, G become coincident in figure 5 and the vertical portions CC, EE in figure 6 coalesce. The limiting value of $\rho = 1/\sqrt{2}$ corresponds to $\Delta'/L = 10^{-2}$, the value which has already been used as the lower limit of the size L for which the models must cater.

Let us now consider how far these postulated mechanisms are likely to be realized in practice. It is curious that although ρ is least in the diametrical plane containing the applied field, so that penetration of induction has been assumed to occur there first, in actual fact the external field is zero in this plane, and is greatest at the sides of the cylinder.

Thus physically we should first expect penetration to occur at the edges, and the experiments of de Haas & Casimir-Jonker (1933) suggest that this might be the case.

A possible mechanism, therefore, is one of penetration at the side (see figure 7 where a schematic picture is given), forming localized normal regions, which as they grow inwards produce regions of low field, causing the reappearance of the superconducting phase at 1, 2 and then 3 in the normal region, the whole branching laminar structure being pushed in sideways. The laminae are not to be considered as necessarily running the whole length of the wire, but it is likely that there will be pockets of growth on alternate sides along the wire. Slight stable penetration at the edges in the manner of (*a*) figure 7, may take place as soon as $h > \frac{1}{2}H_c$, and this would account for the slight deviation from perfect diamagnetism in the interval *KM* (part II, figure 4). Only when it is the stable state, however, will the whole laminar system form, and then fairly suddenly (*MN*, II, figure 4). This point will occur at a value of ρ corresponding to equality of *total* free energies. Since ϕ and $d\phi/dB$ are proportional to $L^{-\frac{1}{2}}$, the appropriate value of L to be used in (14) and (15) is $[\overline{L^{-\frac{1}{2}}}]^{-2} = 1{\cdot}60r$.

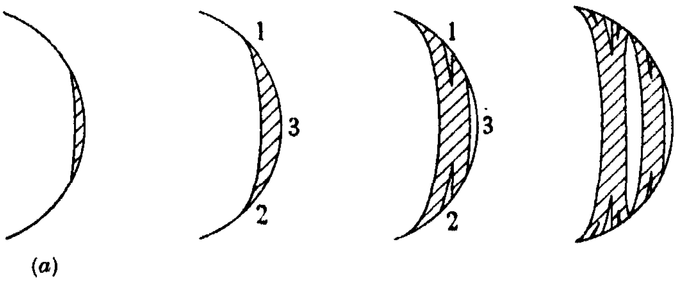


FIGURE 7. Schematic illustration of possible mode of entry into intermediate state.

Relations (14) and (15) thus become:

$$\rho_T = \tfrac{1}{2} + 2{\cdot}31\left(\frac{\Delta'}{r}\right)^{\frac{1}{2}} - 1{\cdot}38\left(\frac{\Delta'}{r}\right)^{\frac{3}{4}}, \tag{20}$$

$$\rho_L = \tfrac{1}{2} + 1{\cdot}283\left(\frac{\Delta'}{r}\right)^{\frac{3}{4}} - 1{\cdot}05\left(\frac{\Delta'}{r}\right)^{\frac{1}{2}}. \tag{21}$$

At the transition from the superconducting state the *I*-*h* relation will still have the same general form of figure 6, as illustrated in figure 8, with some rounding at the top due to the initial penetration just mentioned. The resistance tail *lm* in II, figure 4, would correspond to small local pockets of intermediate state.

Turning to the other end of the curves for the intermediate state, *E*, figure 5 suggests that the normal state becomes the stable state, but one cannot expect the superconducting regions to be suddenly ejected here. It is more likely that they will continue to shrink in a metastable condition. The transition into the normal state will therefore be somewhat as indicated in figure 8, and we note that if the laminar structure, or something very like it, persists until $I = 0$, then this transition occurs at a value of applied field less than H_c, the discrepancy increasing as the cylinder radius r decreases.

It is more difficult to predict the variation of resistance with applied field in the intermediate state, as the geometrical disposition of the normal and superconducting phases has also to be taken into account. If the mean induction B has the value γH_c then the fraction of normal resistance may have any value between 0 and γ. In particular, for a structure of the threads of normal metal embedded in superconducting metal there should be no resistance. However, it is reasonable to expect that if resistance is found, there should be a correlation of the main features of the B-h and resistance-field curves.

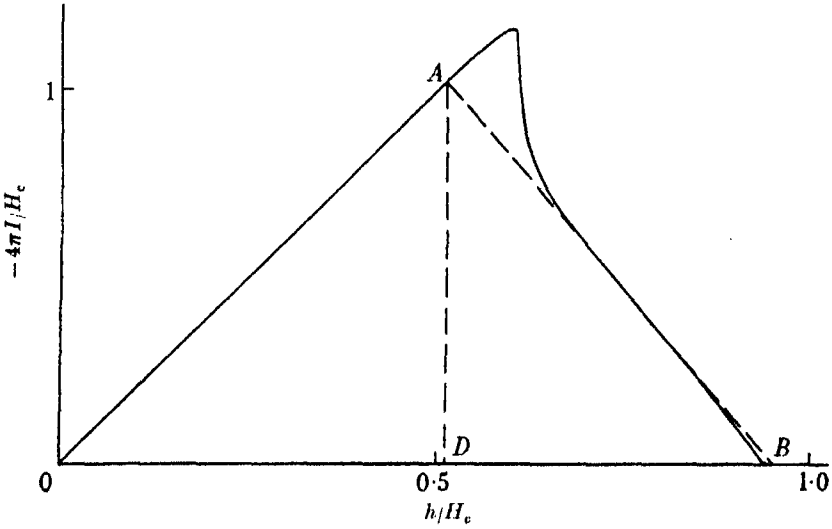


Figure 8. I-h curve for $\Delta'/r = 2{\cdot}4 \times 10^{-3}$.

In connexion with the fitting in I of the theoretical curves to those experimentally found, it is useful to consider the slope dI/dh, of the I-h curve in the intermediate state. For $\Delta'/r = 0$ it is $1/2\pi$, and becomes steeper as Δ'/r increases. Using $I = (B-h)/2\pi$ for a transverse cylinder, in conjunctions with equations (18) and (19), it is found that

$$\frac{dI}{dh} = \frac{1}{2\pi}\frac{1+\epsilon}{1-\epsilon},$$

where $$\epsilon = \frac{1}{0{\cdot}80^{\frac{1}{2}}\,.\,3\,.\,2^{\frac{1}{2}}\,.\,[(\sqrt{2}-1)]^{\frac{1}{2}}\,\gamma^{\frac{1}{2}}(1-\gamma)^{\frac{1}{2}}}\left(\frac{\Delta'}{r}\right)^{\frac{1}{2}} \quad \text{for the laminar model} \tag{22}$$

and $$\epsilon = \frac{\pi^{\frac{1}{2}}(1-\frac{3}{4}\beta)}{0{\cdot}80^{\frac{1}{2}}\,.\,3\,.\,[(\sqrt{2}-1)]^{\frac{1}{2}}\,\gamma^{\frac{1}{2}}(1-\beta)^{\frac{1}{2}}}\left(\frac{\Delta'}{r}\right)^{\frac{1}{2}} \quad \text{for the thread model} \tag{23}$$

with $$\gamma = \frac{B}{H_c} \quad \text{and} \quad \beta = \sqrt{\left(\frac{2B}{\pi H_c}\right)}.$$

If now we construct the triangle OAB, figure 8, from the I-h curve of the superconducting state and any tangent to that of the intermediate state, then the ratio $\sigma = OD/OB = \frac{1}{2}(1+\epsilon)$. Moreover, if we take the tangent at $\gamma = B/H_c = \frac{1}{2}$, we find that ϵ, which is a slowly varying function of γ, has the same value for both models giving

$$\sigma = \tfrac{1}{2} + 2{\cdot}2\left(\frac{\Delta'}{r}\right)^{\frac{1}{2}}. \tag{24}$$

In applying this relation to experimental curves, one must be careful that AB does in fact represent dI/dh for the intermediate state, and is not affected by the transition range from the superconducting state at the upper end, nor by the tails of the transition into the normal state at the lower end.

We now turn to transitions in a decreasing magnetic field. As indicated by E, figure 5, the intermediate state does not become thermodynamically favourable until well below H_c. The cylinder is initially entirely in the normal state, with $I = 0$, and B and h uniform through space. Hence in contrast with the transition from the superconducting state in an increasing field, no portion of the body is more likely than another to be a centre of change, the edges, which played this role in an increasing field, being in no different condition from the remainder. It is likely that the formation of the intermediate state in an increasing field is brought about by a growth of the natural penetration of induction into the surface of a superconductor. There is no counterpart to this in the normal state, and therefore in a decreasing field 'supercooling' effects are to be expected, for just as with the condensation of liquid drops from a supercooled vapour, nuclei of formation are required. Once the transition has begun it proceeds discontinuously, provided that the material is homogeneous. Since BE, figure 5, increases with decrease of cylinder radius, and since smaller cylinders would probably contain less centres of formation of the superconducting phase, the total delay in formation of the intermediate state may be expected to be greater for thinner wires. Just as in an increasing field no sudden expulsion of the superconducting regions was expected, so here in a decreasing field no sudden expulsion of normal regions is expected for $h = \rho H_c$, the intermediate state continuing until the normal laminae or threads have shrunk to nothing. This will occur for a value of h slightly greater than $\frac{1}{2}H_c$.

5. DISCUSSION

On the basis of these models the following main features of the magnetization curves find qualitative explanation:

(*a*) The delay of formation of the intermediate state in an increasing field until $h = \rho H_c$, the ratio ρ having the value $\frac{1}{2}$ for a cylinder of infinite diameter, but increasing as the diameter increases.

(*b*) Slight departure from linearity just before this transition point.

(*c*) Rapid drop of magnetization at this transition point, giving the characteristic 'horn'.

(*d*) In the intermediate state an almost linear I-h relation whose slope dI/dh is greater than the value $1/2\pi$ for an infinite cylinder diameter, by an amount increasing with decrease in diameter.

(*e*) The normal state is reached for $h < H_c$, the difference increasing with decrease in diameter.

(*f*) In a decreasing field, a supercooling which increases with decrease in diameter.

(*g*) In the intermediate state a retracing of the increasing field curve as far as the horn, and continuing almost linearly beyond it, until the curve for the superconducting state is reached.

The combined resistance and magnetization experiment, whose results are illustrated in I, figure 3 and II, figure 4, showed the expected correlation of the main features of both curves. The fractional resistance, was, as suggested, always less than the ratio B/H_c but for sufficiently large measuring currents not very much less, suggesting that the intermediate state realized in practice does have a predominantly laminar character.

The quantitative comparison of theory and experiment, namely the variation of ρ and of dI/dh with the cylinder radius, and the detailed fitting of the magnetization curves, has been shown in I and II to give fairly good but not perfect agreement. We should, however, be cautious in considering these models as giving more than an idealized picture. The shallowness of the free energy minima mentioned in § 2 suggests that a considerable latitude is probable, and the experiments of Meshkovsky & Shalnikov (1947) using a sphere, indicate irregular laminar and thread formation. Since, however, the free energy does not appear to be very dependent on the actual configuration of the intermediate state, these idealized models may be expected to give a reasonable indication of its macroscopic behaviour.

A further result of the experiments of Meshkovsky & Shalnikov, was that although an essentially laminar system was under suitable conditions set up in the intermediate state, the critical width of separation of the two halves of the sphere predicted by Landau (1943), was not found. For quite wide separations the flux left the plane surfaces of the hemispheres with periodicity of the unbranched main laminae. This naturally suggests that the laminae may not branch at all, but come out to all surfaces with the same macroscopic divisions. The results of the experiments proposed by Meshkovsky & Shalnikov to determine whether this is so or not will be of great interest. Examination of the more regular patterns in their paper (figures 11 and 14), shows that the periodicity of the laminae $a_1 + b_1 = 0{\cdot}13 \pm 0{\cdot}02$ cm. Landau (1943) has shown that for a branching laminar system, the periodicity of the principal laminae is given by $a_1 + b_1 \simeq 2\Delta'^{\frac{1}{3}}L^{\frac{2}{3}}$. Taking L as the diameter of the sphere (3·9 cm.), this relation gives the value $\Delta' = 1{\cdot}8 \times 10^{-5}$ cm. at 3° K. The values of Δ' obtained from the magnetization measurements in I and the resistance measurements in II are $1{\cdot}5 \times 10^{-5}$ and $1{\cdot}8 \times 10^{-5}$ cm. respectively at this temperature (see I, table 2). The agreement may be accidental, but it is possible that even without branching the periodicity of the laminae may be given by a relation of the form just quoted. The effect that lack of branching would have on the present theory would be to remove the last term in the brackets of equation (6). The free energy f would then have no minimum, very thick laminae being favoured. In practice the periodicity of the laminae would probably then be determined by conditions at the outside surface of the metal, and by the mechanism and rate of formation.

I gladly record the benefits I have received from discussions with Dr D. Shoenberg and Mr A. B. Pippard. Most of the work was carried out during the tenure of an award from the Department of Scientific and Industrial Research.

APPENDIX

A. *Detailed calculations of the thread model*

Step (i) described in § 2 yields relation (1) directly.

Step (ii) proceeds exactly as for the laminar model, the results (2) for the two models differing by a factor 2 under the square-root sign. The factor occurs because a cylinder has twice the surface area/volume ratio of a plate having a thickness equal to the cylinder diameter.

The relation between the angles θ_n, (3), follows from (1) and (2), and is the same for both models. The reproduction of the arrays of figure 2 half-size (step (iii)) requires $x_1\theta_1 = 2^{n-1}x_n\theta_n$ as for the laminar model, and hence the relation between the distances x_n of branching processes from the surface, is also the same for the two models (relation (4)). Step (iii) also requires that the maximum deflexion of an infinitely subdivided thread shall be half the distance between the surfaces of two diagonally arranged principal threads, as may be seen from an examination of figure 2. This gives equation (5) of table 1:

$$\sum_1^{\infty} x_n\theta_n = 2x_1\theta_1 = \frac{b_1}{\sqrt{2}} - \frac{a_1}{2}.$$

Step (iv). The mean value of B is evidently given by $\frac{1}{4}\pi a_1^2 H_c = b_1^2 B$ which yields (6).

Step (v). The free energy of the system is now formed, reckoning that of the superconducting state as zero. If the system just consisted of all the main threads of diameter a_1 running right through without branching, then the free energy per unit area of plate would be:

$$\frac{1}{b_1^2}\left\{\tfrac{1}{4}\pi a_1^2 L \frac{H_c^2}{4\pi} + \pi a_1 L \frac{H_c^2}{8\pi}\Delta\right\}. \tag{A 1}$$

If now the threads branch once only (at each end), at distance x_1 from the surface there is additional surface energy: $\pi(4a_2 - a_1)\,2x_1\frac{H_c^2}{8\pi}\Delta$. There is also an additional volume energy, which, as a consequence of step (ii), has the same value. Extending the argument to the complete model with an infinity of branching processes, the additional free energy is: $2\frac{H_c^2}{8\pi}\Delta\,.\,\pi\sum_1^{\infty}[4^{n-1}(4a_{n+1}-a_n)\,2x_n]$. Performing this summation with the aid of (1) and (4), and adding the free energy of the unbranched threads (A 1), in which a substitution for b_1 in the first term is made from (6), we arrive at the free energy per unit area of the

$$f = \frac{H_c B L}{4\pi} + \frac{H_c^2 a_1 \Delta}{8b_1^2}\left\{L + \frac{4x_1}{1-1/\sqrt{2}}\right\}. \tag{A 2}$$

From (5), using (2) and (6) we have $x_1 = \frac{a_1^{\frac{3}{2}}}{8}\sqrt{\left(\frac{\pi H_c}{2\Delta B}\right)}(1-\beta)$, where $\beta = \sqrt{\left(\frac{2B}{\pi H_c}\right)}$. Substitution of this value of x_1 in (A 2) yields (7), table 1, where f is now a function of one variable a_1 only. After minimization of f with respect to a_1, the free energy per unit volume $\Phi/4\pi$ has the form (8), the minimum occurring at a value of a_1 given by (9). By putting $\Phi = H_c B + \phi$ the free energy arising from the surface energy is separated off as $\phi/4\pi$ per unit volume.

B. *The point of entry into the intermediate state*

Consider first the thread model. Substitution for ϕ and $d\phi/dB$ from table 1 in (12) gives after rearrangement:

$$\frac{B}{H_c} = \left[\frac{2\sqrt{\pi}}{(\sqrt{2}-1)}\frac{\Delta'}{L}\right]^{\frac{1}{4}}(1-\beta)^{-\frac{1}{4}}. \tag{A 3}$$

This value of B is substituted in $d\phi/dB$, and both are then put in (13). After expanding h_t as a power series in β, and substituting for B from (A 3) to whatever degree of approximation is required, the ratio $\rho_T = h_t/H_c$ is found as a power series in $(\Delta'/L)^{\frac{1}{4}}$, of which (14) shows the important terms. A similar procedure for the laminar model gives the corresponding ratio ρ_L as a power series in $(\Delta'/L)^{\frac{1}{2}}$ the initial terms being given in (15). Numerical solution of (13) for particular values of Δ'/L shows that (14) and (15) are accurate to about 0·1 % for $\Delta'/L = 10^{-3}$, and within 1 % for $\Delta'/L = 10^{-2}$.

REFERENCES

Ginsburg, V. 1945 *J. Phys. U.S.S.R.* **9**, 305.
de Haas, W. J. & Casimir-Jonker, J. M. 1933 *Physica*, **1**, 291.
Landau, L. 1943 *J. Phys. U.S.S.R.* **7**, 99.
London, H. 1935 *Proc. Roy. Soc.* A, **152**, 650.
Meshkovsky, A. & Shalnikov, A. 1947 *J. Phys. U.S.S.R.* **11**, 1.

Aspects of a theory of unimolecular reaction rates

By N. B. Slater, *University of Leeds*

(*Communicated by M. G. Evans, F.R.S.—Received* 29 *December* 1947)

A uniform gas of polyatomic molecules is treated as an assembly of classical vibrating systems, which dissociate when one internal co-ordinate q reaches a critically high value q_0 which is related to the dissociation energy E_0. The unimolecular velocity constant at temperature T is found to be

$$K = \nu \exp(-E_0/kT),$$

where the 'frequency factor' ν lies in the range of molecular vibration frequencies. The factor ν may be interpreted (i) as a weighted average of the normal vibration frequencies, or (ii) as the ratio of the product of the normal frequencies to the product of the frequencies with q fixed, or (iii) in the case where dissociation is due to the rupture of an isolated bond, as the vibration frequency of an imaginary diatomic molecule consisting merely of the two bonded atoms connected by the original bond-force.

The dissociation rate is formulated in several ways, which represent different aspects of the physical picture. The main contrast is between (*a*) the rate as the average frequency with which the normal-mode vibrations come sufficiently into phase to carry q to the critical value q_0, and (*b*) 'transition state' formulations, giving the rate as the product of the relative concentration of activated complexes (molecules with q near q_0) and the mean transition frequency. The mathematical equivalence of these methods is shown by a study of the asymptotic distribution of values of sums of harmonic vibrations.

The present model is used to illustrate some concepts of transition state or activated complex theory, such as the 'effective mass' in the reaction co-ordinate, and the partition function of the activated complex. The relation of the model to Kassel's theory is shown by calculating the dissociation rates of molecules of specified total energy.

1. INTRODUCTION

This theory of unimolecular reaction rates in gases is based on the model of a polyatomic molecule as a classical vibrating system which dissociates when one particular internal co-ordinate, such as a bond-extension, attains a critically high value. The general formula derived for the high-pressure rate at absolute temperature T is

$$-\frac{1}{N}\frac{dN}{dt} \equiv K = \nu \exp(-E_0/kT)\,\mathrm{sec.}^{-2}, \tag{1}$$

where N is the concentration, k is Boltzmann's constant and E_0 the 'activation energy' per molecule. The frequency factor ν lies between the greatest and least of the vibration frequencies of the molecule.

Polanyi & Wigner (1928) and Pelzer (1933) obtained reaction rates of this nature for simple chain molecules. For the general molecule approximate, and later precise, formulae of this type have since been derived (Slater 1939, 1947*a*). Now the model giving this result is crude, in that it uses classical mechanics and ignores anharmonicity of the vibrations. Its simplicity, however, makes it suitable for comparing different approaches to rate problems, and also for illustrating some concepts of transition state theory. Using the same model of dissociation we shall therefore derive the rate K by several methods, and examine their relation to each other and to the general formalism of the transition state as expounded by Glasstone, Laidler & Eyring (1941). We summarize in (i) to (vi) below these different approaches, and then state the corresponding forms of ν in the rate (1).

(i) In § 3 the internal co-ordinate q which is to break at a critical value q_0 is treated as a sum of normal-mode vibrations:

$$q \equiv F(t) = \sum_{s=1}^{n} \alpha_s \sqrt{\epsilon_s} \cos 2\pi(\nu_s t + \psi_s), \tag{2}$$

where ϵ_s and ν_s are the energy and frequency in the sth mode. A formula of Kac (1943) gives the asymptotic frequency, $2L$, of zeros of $F(t) - q_0$ in the case where the ν_s are linearly independent numbers. The rate K, (1), is obtained as the average of L over an equilibrium distribution of the energies ϵ_s. This formulation, and the significance of linear independence, is discussed in § 4.

(ii) In § 5 it is shown that the frequency $2L$ can be expressed in terms of $m(q, \dot{q})$, the distribution function of a sum q, (2), and its derivative $\dot{q}$. This suggests the formulation of K in § 6 as the mean of $\dot{q}$, the velocity in the reaction co-ordinate, over the distribution $m(q_0, \dot{q})$ averaged over energy states.

(iii) In § 7, an extension of Pelzer's method, K is formulated as the direct average of $\dot{q}$ over an equilibrium distribution in phase-space, where the kinetic and potential energies are the general quadratic forms

$$T' = \tfrac{1}{2}\Sigma\Sigma a_{rs}\dot{q}_r\dot{q}_s, \quad V = \tfrac{1}{2}\Sigma\Sigma b_{rs}q_r q_s \quad (q \equiv q_1). \tag{3}$$

No use is made here of normal modes, but in § 8 the equivalence of this method to that of (i) is shown by a transformation of phase-space.

(iv) In transition state theory, a rate is calculated as the product of the population of 'activated complexes' (in our case, molecules with q near q_0) and of the mean transition velocity through the critical configuration. These concepts have been used in (ii) and (iii), but in § 9 the rate is formulated in direct imitation of Eyring's partition function treatment. The physical significance of the transition velocity and of the partition function of the activated complex is discussed.

(v) In § 10 the model is illustrated by the potential energy diagrams used in transition state theory, with particular reference to the error of the assumption that the vibrations are exactly harmonic.

(vi) Finally in § 11 the dissociation rate is derived for molecules of specified total energy. The result is formally identical with the specific rate derived by Kassel (1932) on a different model of the dissociation process.

The constant ν takes different forms in the derivations (i) to (iv) of the rate (1), namely

$$\left.\begin{aligned} &\nu = \left(\frac{\Sigma\alpha_s^2\nu_s^2}{\Sigma\alpha_s^2}\right)^{\frac{1}{2}} \quad \text{(i), (ii);} \qquad \nu = \frac{1}{2\pi}\left(\frac{A_{11}B}{AB_{11}}\right)^{\frac{1}{2}} \quad \text{(iii);} \\ &\qquad\qquad \nu = \nu_1\nu_2\dots\nu_n/(\nu_2'\dots\nu_n'). \qquad\qquad \text{(iv).} \end{aligned}\right\} \tag{4}$$

Here ν_s, α_s are as in (2): A, B denote the determinants $\|a_{rs}\|$, $\|b_{rs}\|$ of (3) and A_{11}, B_{11} the co-factors of a_{11}, b_{11}: $\nu_2', \dots, \nu_n'$ are the frequencies obtained with a fixed value of q in (3), and correspond to the 'frequencies of the activated complex'. The equivalence of the forms (4) will be shown first in § 2. In the important case where dissociation is due to the breaking of an isolated bond between two atoms, ν is equal to the vibration frequency of an imaginary diatomic molecule consisting of these two atoms connected by the same bond-force as in the polyatomic molecule.

2. MOLECULAR VIBRATIONS

Some formulae are derived here concerning the normal modes of a vibrating system, and the 'average frequency' ν which will occur in the dissociation rate.

The internal kinetic energy T' (marked to distinguish it from the temperature T) and potential energy V of a polyatomic molecule will be taken to be the classical positive definite quadratic forms:

$$T' = \tfrac{1}{2}\sum_1^n\sum a_{rs}\dot{q}_r\dot{q}_s, \quad V = \tfrac{1}{2}\sum_1^n\sum b_{rs}q_rq_s. \tag{5}$$

Here n is the number of internal degrees of freedom; rotation of the whole molecule is ignored. Typical co-ordinates q_r are elongations of interatomic bonds, or changes in bond angles, from the equilibrium configuration; the inertia coefficients a_{rs} then depend on the atomic masses and equilibrium bond-lengths. By treating the 'force constants' b_{rs} as fixed we neglect anharmonicity of the vibrations.

Let $Q_1, \dots, Q_n$ be normal co-ordinates, in terms of which

$$T' = \sum_1^n \dot{Q}_s^2/\lambda_s^2, \quad V = \sum_1^n Q_s^2 \quad (\lambda_s = 2\pi\nu_s), \tag{6}$$

where the normal mode frequencies $\nu_1, \dots, \nu_n$ will be assumed to be distinct. The solution of the motion is, from (6),

$$Q_s = \sqrt{\epsilon_s}\cos 2\pi(\nu_s t + \psi_s) \quad (s = 1, \dots, n),$$

where ϵ_s is the energy in the sth normal mode and ψ_s is a phase constant. If $\boldsymbol{\alpha} = (\alpha_{rs})$ is the matrix of transformation from q to Q, so that

$$q_r = \sum_s \alpha_{rs} Q_s, \tag{7}$$

then the solution for one co-ordinate, for example q_1, is

$$q_1 = \sum_s \alpha_{1s} \sqrt{\epsilon_s} \cos 2\pi(\nu_s t + \psi_s). \tag{8}$$

We shall be interested mainly in q_1, and shall usually omit the suffices 1 in (8). We refer to the α_{1s} or α_s as the 'amplitude factors' for the co-ordinate q_1.

The '*average frequency*' ν associated with co-ordinate q_1 will be defined first by

$$\nu = \lambda/2\pi, \quad \lambda^2 = \alpha^{-2} \sum_1^n \alpha_s^2 \lambda_s^2, \tag{9}$$

where

$$\alpha^2 = \sum_1^n \alpha_s^2. \tag{10}$$

Thus ν^2 is the mean of the ν_s^2, weighted by the amplitude factors α_s^2 of co-ordinate q_1. We shall now derive two alternative formulae for ν.

Since (7) reduces the forms (5) to (6), in matrix notation

$$\tfrac{1}{2}\tilde{\boldsymbol{\alpha}}\mathbf{a}\boldsymbol{\alpha} = \boldsymbol{\Lambda}^{-1}, \quad \tfrac{1}{2}\tilde{\boldsymbol{\alpha}}\mathbf{b}\boldsymbol{\alpha} = \mathbf{1},$$

where $\tilde{\boldsymbol{\alpha}}$ is the transpose of $\boldsymbol{\alpha}$, and $\boldsymbol{\Lambda}$ is the diagonal matrix with elements $\lambda_1^2, \ldots, \lambda_n^2$. Multiplying these equations each by initial $\tilde{\boldsymbol{\alpha}}^{-1}$ and final $\boldsymbol{\alpha}^{-1}$ and inverting the results, we obtain

$$\boldsymbol{\alpha}\boldsymbol{\Lambda}\tilde{\boldsymbol{\alpha}} = 2\mathbf{a}^{-1}, \quad \boldsymbol{\alpha}\tilde{\boldsymbol{\alpha}} = 2\mathbf{b}^{-1}.$$

The leading elements of these matrix equations are

$$\Sigma \alpha_{1s}^2 \lambda_s^2 = 2A_{11}/A, \quad \Sigma \alpha_{1s}^2 = 2B_{11}/B,$$

where A, B denote the determinants $\|a_{rs}\|$, $\|b_{rs}\|$ and A_{11}, B_{11} the co-factors of a_{11}, b_{11}. Hence in (9), (10)

$$\alpha^2\lambda^2 = 2A_{11}/A, \quad \alpha^2 = 2B_{11}/B, \tag{11}$$

and

$$\nu = \frac{\lambda}{2\pi} = \frac{1}{2\pi}\left(\frac{A_{11}B}{AB_{11}}\right)^{\frac{1}{2}}, \tag{12}$$

which is the second form of ν.

To obtain a third form, we note that

$$B/A = \lambda_1^2 \lambda_2^2 \ldots \lambda_n^2,$$

since the λ_s^2 are the roots in λ^2 of the secular equation

$$\|a_{rs}\lambda^2 - b_{rs}\| = 0 \quad (r, s = 1, \ldots, n). \tag{13}$$

Let $\lambda_2'^2, \ldots, \lambda_n'^2$ be the roots of the equation obtained by omitting the first row and column of (13): then similarly

$$B_{11}/A_{11} = \lambda_2'^2 \ldots \lambda_n'^2. \tag{14}$$

Hence (12) may be written

$$\nu = \nu_1 \nu_2 \dots \nu_n/(\nu_2' \dots \nu_n') \quad (\nu_s' \equiv \lambda_s'/2\pi), \tag{15}$$

which is the third form. The ν_s' are the normal frequencies for the system constrained to have q_1 fixed at zero (or at any other value), and so by a well-known theorem their magnitudes are interlaced with those of the original frequencies ν_s. It follows then from (15), or directly from the first definition (9), that ν lies between the greatest and least of the ν_s.

The 'activation energy'. By (8), q can just reach a prescribed positive value q_0 if the energies ϵ_s satisfy

$$\Sigma \,|\, \alpha_s \,|\, \sqrt{\epsilon_s} = q_0.$$

The values ϵ_s satisfying this and making the total energy $E = \Sigma\epsilon_s$ a minimum are easily found to be

$$\epsilon_s = \epsilon_{s0} = q_0^2 \alpha_s^2 \alpha^{-4}, \tag{16}$$

where α^2 is defined by (10). Hence the minimum total energy E_0 for q to attain q_0 is

$$E_0 = \Sigma\epsilon_{s0} = q_0^2/\alpha^2. \tag{17}$$

This may also be written, by (11),

$$E_0 = \frac{1}{2}\frac{B}{B_{11}} q_0^2. \tag{18}$$

We shall later identify E_0 with the 'activation energy'.

A particular case. If the co-ordinate q_1 is the elongation of the distance between two atoms of masses m_1, m_2, then it can be shown that, irrespective of the nature of the remaining co-ordinates $q_2 \dots q_n$,

$$A/A_{11} = m_1 m_2/(m_1 + m_2) \equiv m, \tag{19}$$

which is the 'reduced mass' of the two atoms in relative motion. If, further, the potential energy contains no cross terms $q_1 q_s$, so that in (5)

$$2V = b_{11} q_1^2 + \sum_{2}^{n}\sum b_{rs} q_r q_s, \tag{20}$$

then

$$B/B_{11} = b_{11}. \tag{21}$$

Hence in this case, by (12), (18),

$$\nu = \frac{1}{2\pi}\sqrt{\frac{b_{11}}{m}}, \quad E_0 = \tfrac{1}{2} b_{11} q_0^2. \tag{22}$$

Thus ν is here the vibration frequency of an imaginary diatomic molecule with masses m_1, m_2 and force constant b_{11}, and E_0 is its energy at extension q_0.

3. THE UNIMOLECULAR VELOCITY CONSTANT

In this section the dissociation rate is formulated, and evaluated by using a formula of Kac for the frequency of zeros of a trigonometrical sum. The validity of the formulation will be discussed in § 4, and Kac's formula in § 5.

We consider the dissociation at high pressure of a gas of polyatomic molecules represented by the vibrating systems treated in § 2. We assume that a molecule dissociates when a particular internal co-ordinate $q \equiv q_1$ attains a large critical value q_0; we assume also that the vibrations remain harmonic up to the point of rupture. In an undissociated molecule, the behaviour of q is given by (8), namely

$$q \equiv F(t) = \sum_1^n a_s \cos 2\pi(\nu_s t + \psi_s), \tag{23}$$

with

$$a_s = \alpha_s \sqrt{\epsilon_s}. \tag{24}$$

The energies ϵ_s and phase constants ψ_s change only when the molecule collides. The molecule is said to be *energized* if q can reach q_0, that is if the ϵ_s satisfy

$$\Sigma\,|\,a_s\,| \equiv \Sigma\,|\,\alpha_s\,|\,\sqrt{\epsilon_s} \geqq q_0. \tag{25}$$

A molecule energized by collision will not dissociate, however, unless the terms in (23) come sufficiently into phase for q to reach q_0 before further collision removes the high energy. If, as we assume, the pressure is large enough, collisions will be much more frequent than the attainment of the dissociation configuration, and so will effectively maintain a steady population of energized molecules which is slowly depleted by dissociation. We assume, then, that

(*a*) the 'dissociation frequency', or fraction dissociating per second, of energized molecules with specified values $(\epsilon_1, \ldots, \epsilon_n)$ is equal to L, the average frequency with which $F(t)$, (23), rises to the value q_0; and

(*b*) the number of molecules, out of a total concentration N, which have energies in the ranges $(\epsilon_1, \epsilon_1 + d\epsilon_1) \ldots (\epsilon_n, \epsilon_n + d\epsilon_n)$ is effectively the equilibrium number

$$Nf(\epsilon)\,d\epsilon_1 \ldots d\epsilon_n, \quad f(\epsilon) \equiv (kT)^{-n} \exp(-\Sigma\epsilon_s/kT), \tag{26}$$

where T is the absolute temperature and k is Boltzmann's constant.

These assumptions lead immediately to the dissociation rate; for as a fraction L of the molecules (26) in any small energized range dissociate in unit time, the overall unimolecular velocity constant is

$$K \equiv -\frac{1}{N}\frac{dN}{dt} = \int \ldots \int Lf(\epsilon)\,d\epsilon_1 \ldots d\epsilon_n, \tag{27}$$

where the integral is over the ranges $\epsilon_s > 0$ satisfying (25).

Two types of definition may be suggested for the dissociation frequency L, namely a long-time average L_1 and a phase-average L_2:

$$L_1 = \tfrac{1}{2} \lim_{\tau \to \infty} \{M(\tau)/\tau\}, \tag{28}$$

$$L_2 = \frac{1}{2\tau} \int_0^1 \ldots \int_0^1 M(\tau)\,d\psi_1 \ldots d\psi_n, \tag{29}$$

where $M(\tau)$ is the number of zeros of $F(t) - q_0$ [(23)] in the time $0 \leqq t < \tau$. (The factor $\frac{1}{2}$ represents our selection of zeros for $F(t)$ rising to q_0.) We shall assume that the frequencies $\nu_1, \ldots, \nu_n$ in (23) are *linearly independent* in the field of rational numbers, that is that there is no set of integers i_s, not all zero, such that

$$\sum_1^n i_s \nu_s = 0. \qquad (30$$

In this case L_1 and L_2 are equal, and have the value (Kac 1943)

$$L = \frac{1}{4\pi^2}\iint_{-\infty}^{\infty} \frac{\cos(q_0\xi)}{\eta^2}\left\{\prod_1^n J_0(a_s\xi) - \prod_1^n J_0[a_s\sqrt{(\xi^2+\eta^2\lambda_s^2)}]\right\} d\eta\, d\xi. \tag{31}$$

We use this to evaluate (27). As L vanishes where (25) is not satisfied, each ϵ_s in (27) may be given the range $(0, \infty)$. By an inversion of the order of integration, (27) then becomes

$$K = \frac{1}{4\pi^2}\iint_{-\infty}^{\infty} \frac{\cos(q_0\xi)}{\eta^2}\left\{\prod_1^n I_s(\xi, 0) - \prod_1^n I_s(\xi, \eta)\right\} d\eta\, d\xi,$$

where

$$\begin{aligned} I_s(\xi,\eta) &\equiv \int_0^\infty J_0[\alpha_s\sqrt{\{\epsilon_s(\xi^2+\eta^2\lambda_s^2)\}}]\exp(-\epsilon_s/kT)\,d\epsilon_s/kT \\ &= \exp\{-\tfrac{1}{4}kT\alpha_s^2(\xi^2+\eta^2\lambda_s^2)\}, \end{aligned} \tag{32}$$

by the use of Weber's first exponential formula (Watson 1922, p. 393). Thus by the definitions (9), (10) of λ, ν and α,

$$\begin{aligned} K &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\cos(q_0\xi)\exp(-\tfrac{1}{4}kT\alpha^2\xi^2)\,d\xi\int_{-\infty}^{\infty}[1-\exp(-\tfrac{1}{4}kT\alpha^2\lambda^2\eta^2)]\frac{d\eta}{\eta^2} \\ &= \nu\exp(-q_0^2/\alpha^2kT), \end{aligned} \tag{33}$$

by well-known formulae. In terms of the energy E_0 of (17), the result takes the final form

$$K = \nu\exp(-E_0/kT). \tag{34}$$

Here E_0 is the minimum energy for dissociation (that is, for q to reach q_0), and for comparison with experimental rates is to be identified with the activation energy per molecule. As we mentioned in § 2, ν lies in the range of the molecular vibration frequencies, and so is in effect confined to the range 10^{12} to 10^{14} sec.$^{-1}$. Its value may be calculated from (12), or from (22) in the case where dissociation is due to the rupture of an isolated bond.

4. NOTES ON THE FORMULATION

We examine the significance of some assumptions made in § 3. First we estimate the relative magnitudes of the dissociation frequency L and of collision and vibration frequencies. Secondly we shall discuss the condition of linear independence of the vibration frequencies, and the correct definition of L.

For the distribution (26) of energies to be effectively maintained, it was assumed that L is small compared with $\nu_{\text{coll.}}$, the frequency of collisions per molecule. We take the simple case of a molecule for which the amplitude factors α_s (24) are approximately the same, and likewise the normal frequencies ν_s. If $E_0 > nkT$, the average molecule with energy greater than E_0 will have total energy approximately $E_0 + kT$. The distribution most favourable to dissociation will be approximate equipartition:

$$\epsilon_s \simeq (E_0 + kT)/n \quad (s = 1, \ldots, n). \tag{35}$$

Using my previous approximate formula (Slater 1939, equation (8)) for L, and the typical values

$$E_0/kT = 40, \quad \nu = 10^{13}\,\text{sec.}^{-1},$$

we find for the energies (35) that

when	$n = 1$	5	10	15,
then	$L = 10^{13}$	10^9	10^4	$10^{-1}\,\text{sec.}^{-1}$.

If the energy distribution is more uneven than in (35), then L will be smaller.

Collision frequencies per molecule, as typified by ethane at atmospheric pressure and temperature 500° K, are of the order

$$\nu_{\text{coll.}} = 10^9\,\text{sec.}^{-1}.$$

Thus for polyatomic molecules, with $n \geqq 10$, at normal pressures

$$L \ll \nu_{\text{coll.}} \ll \nu, \tag{36}$$

so that the assumption that L is small compared with ν_{coll} is valid. For simple molecules with small n, $\nu_{\text{coll.}}$ may greatly exceed L only at impossibly high pressures, so that a first-order reaction would not be observed: this agrees with experimental results.

Linear independence and the dissociation frequency

When the vibration frequencies ν_s are not linearly independent the limit L_1 of (28) depends on the phase constants ψ_s in (23), although the phase-average (29) still has the value (31). If the ψ_s were after a collision randomly distributed over the range (0, 1), then an average of the type (29) would correctly represent the dissociation frequency in all cases. Energizing collisions tend, however, to concentrate the phases towards the values $\frac{1}{4}$ or $\frac{3}{4}$, just as a pendulum suddenly struck from small into large oscillation begins its new large swing relatively near the middle. Thus the initial phases are not entirely random. As, however, the inequalities (36) indicate, there is time after an energizing collision for many individual oscillations (of average frequency ν) to occur and smooth out any initial phase-concentrations. These concentrations will recur quickly if the frequencies obey relations of the type (30) with small i_s, but not if the frequencies are linearly independent. This indicates that linear independence is of physical significance in the present classical treatment. It will be remembered that independence is required in quantal theory if the normal modes are to be quantized separately.

A deeper analysis of the 'dissociation frequency' would require consideration of the evenness of spacing of zeros of $F(t)-q_0$, and of the effect of an initial concentration of phase on the time to the next zero. This has not been attempted here: a problem of this type has been investigated by Rice (1945).

5. THE DISTRIBUTION OF VALUES OF A TRIGONOMETRICAL SUM

In this mathematical interlude we shall sketch proofs of the basic formulae used in §§ 3 and 6. We derive first the asymptotic distributions of a sum $F(t)$, (23), representing the co-ordinate q, and of dF/dt representing the velocity $\dot{q}$. We then express $2L_1$, (28), the asymptotic frequency of zeros of $F(t)-q_0$, as an integral of $\dot{q}$ over the distribution of q and $\dot{q}$, and indicate how this leads to Kac's formula (31).

Let θ_s denote the fractional part of $\nu_s t + \psi_s$, so that

$$0 \leq \theta_s < 1 \quad (s = 1, \ldots, n). \tag{37}$$

We employ the following form of the Kronecker-Weyl approximation theorem: if $\nu_1, \ldots, \nu_n$ are linearly independent, and $(\theta_1, \ldots, \theta_n)$ are considered as the rectangular co-ordinates of a point P_t in the 'unit cube' (37); and if '$\text{meas}_\tau(P_t \text{ in } \Theta)$' denotes the measure of t, out of $0 \leqq t < \tau$, for which P_t lies in a region Θ within the unit cube, then

$$\lim_{\tau\to\infty}\left\{\frac{1}{\tau}\text{meas}_\tau(P_t \text{ in } \Theta)\right\} = \text{extension of } \Theta. \tag{38}$$

Consider first the sum $F(t)$ of (23), which in terms of the θ_s may be written

$$F = \sum_1^n a_s \cos 2\pi\theta_s, \tag{39}$$

and let
$$m(q_0)\,dq_0 \equiv \lim_{\tau\to\infty}\left\{\frac{1}{\tau}\text{meas}_\tau\,(q_0 < F(t) < q_0 + dq_0)\right\} \tag{40}$$

be the 'fraction of time' for which $F(t)$ lies in a short range dq_0. By the theorem (38)

$$m(q_0)\,dq_0 = \int\ldots\int d\theta_1 \ldots d\theta_n, \tag{41}$$

where the integral is over values in (37) such that F (the function (39) with $\theta_1, \ldots, \theta_n$ now treated as independent variables) lies in $(q_0, q_0 + dq_0)$. By (41)

$$\begin{aligned}\int_{-\infty}^{\infty} m(q_0)\exp(iq_0\xi)\,dq_0 &= \int_0^1 \ldots \int_0^1 \exp(iF\xi)\,d\theta_1 \ldots d\theta_n \\ &= \prod_1^n J_0(a_s\xi),\end{aligned} \tag{42}$$

by a formula for the Bessel function J_0. The Fourier transform of (42) is

$$\begin{aligned} m(q_0) &= (2\pi)^{-1}\int_{-\infty}^{\infty} \exp(-iq_0\xi)\,\Pi\,J_0(a_s\xi)\,d\xi \\ &= (2\pi)^{-1}\int_{-\infty}^{\infty} \cos(q_0\xi)\,\Pi\,J_0(a_s\xi)\,d\xi. \end{aligned} \tag{43}$$

This result could also be inferred from a discussion by Wintner (1932). The present method is easily extended to give the distribution of $F(t)$ and of the derivative

$$F'(t) = -\Sigma a_s \lambda_s \sin 2\pi(\nu_s t + \psi_s) \quad (\lambda_s = 2\pi\nu_s),$$

which in terms of the θ_s can be written

$$F' = -\Sigma a_s \lambda_s \sin 2\pi\theta_s. \tag{44}$$

For, if the simultaneous distribution of F and F' is given by

$$m(q_0, \dot q)\,dq_0 d\dot q = \lim_{\tau\to\infty}\left\{\frac{1}{\tau}\text{meas}_\tau\,[q_0 < F(t) < q_0 + dq_0, \dot q < F'(t) < \dot q + d\dot q]\right\}, \tag{45}$$

then by a simple extension we find

$$m(q_0, \dot q) = (2\pi)^{-2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^1 \ldots \int_0^1 \exp\{i(F - q_0)\xi + i(F' - \dot q)\eta\}\,d\theta_1 \ldots d\theta_n\,d\eta\,d\xi, \tag{46}$$

where F, F' denote (39), (44) as functions of the θ_s. Hence

$$m(q_0, \dot{q}) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cos(q_0\xi)\cos(\dot{q}\eta)\prod_1^n J_0[a_s\sqrt{(\xi^2+\eta^2\lambda_s^2)}]\,d\eta\,d\xi. \tag{47}$$

For a given value q_0 of F, the possible values $\dot{q}$ of F' are limited to a symmetrical range $(-U, U)$ depending on q_0. Thus

$$m(q_0, \dot{q}) = 0 \quad \text{for} \quad |\dot{q}| > U. \tag{48}$$

It is to be inferred from the definitions that

$$\int_{-\infty}^{\infty} m(q_0, \dot{q})\,d\dot{q} = \int_{-U}^{U} m(q_0, \dot{q})\,d\dot{q} = m(q_0). \tag{49}$$

The frequency of zeros of $F(t) - q_0$

Out of a long time τ, the time for which $F(t)$ and $F'(t)$ lie simultaneously in specified ranges dq_0, $d\dot{q}$ is by (45) asymptotically equal to

$$\tau m(q_0, \dot{q})\,dq_0\,d\dot{q}.$$

The time taken by $F(t)$ to pass through the range $(q_0, q_0 + dq_0)$ with $F' = \dot{q}$ is

$$dq_0/|\dot{q}|.$$

The ratio, $\tau m(q_0, \dot{q})\,|\dot{q}|\,d\dot{q}$, of these expressions therefore gives the number of 'passages', that is the number of zeros of $F(t) - q_0$, with a given value $\dot{q}$ of F'. The total number of zeros in time τ is therefore

$$M(\tau) \sim \tau \int m(q_0, \dot{q})\,|\dot{q}|\,d\dot{q},$$

where by (48) the limits of integration (here and in the following integrals of $d\dot{q}$) are $(-U, U)$ or $(-\infty, \infty)$. Hence the asymptotic frequency of zeros (28) is

$$2L_1 = \lim_{\tau\to\infty}\frac{M(\tau)}{\tau} = \int m(q_0, \dot{q})\,|\dot{q}|\,d\dot{q}. \tag{50}$$

It is worth noting that if v_0 denotes the average of $|F'|$ for a given value q_0 of F, so that

$$v_0 = \int m(q_0, \dot{q})\,|\dot{q}|\,d\dot{q} \Big/ \int m(q_0, \dot{q})\,d\dot{q},$$

then by (49) we may write (50) as

$$2L_1 = m(q_0)\,v_0, \tag{51}$$

which expresses L_1 in a manner typical of 'transition state' theory.

To derive Kac's formula (31), we insert (46) in (50) and integrate first for $\dot{q}$ and η, obtaining

$$2L_1 = (2\pi)^{-1}\int_{-\infty}^{\infty}\int_0^1 \dots \int_0^1 \exp\{i(F - q_0)\xi\}\,|F'|\,d\theta_1 \dots d\theta_n\,d\xi. \tag{52}$$

To proceed, we write

$$|F'| = \pi^{-1}\int_{-\infty}^{\infty} \eta^{-2}[1 - \cos(F'\eta)]\,d\eta,$$

and thence derive (31) by a process similar to the last stages of Kac's derivation (Kac 1943).

6. THE VELOCITY CONSTANT IN TERMS OF DISTRIBUTION FUNCTIONS

The argument leading to the formula (50) for L_1 suggests an analogous formulation of the velocity constant K. We again assume that the frequencies ν_s are linearly independent. The fraction of molecules with energies $(\epsilon_1, \ldots, \epsilon_n)$ which have the breaking co-ordinate q in the critical region $(q_0 - dq_0, q_0)$ with also $\dot{q}$ in a range $(\dot{q}, \dot{q}+d\dot{q})$ may be represented by (45), where $q \equiv F(t)$ is given by (23), (24). Hence for an equilibrium distribution (26), the number of molecules of all energies with q, $\dot{q}$ in these ranges is $NM(q_0, \dot{q})\,dq_0 d\dot{q}$, where

$$M(q_0, \dot{q}) = \int_0^\infty \ldots \int_0^\infty m(q_0, \dot{q}) f(\epsilon)\, d\epsilon_1 \ldots d\epsilon_n. \tag{53}$$

Inserting (26) and (47) for f and $m(q_0, \dot{q})$, we obtain

$$\begin{aligned} M(q_0, \dot{q}) &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \cos(q_0 \xi) \cos(\dot{q}\eta) \prod_1^n I_s(\xi, \eta)\, d\eta\, d\xi \\ &= \frac{1}{\pi\alpha^2\lambda kT} \exp\left(-\frac{E_0}{kT} - \frac{\dot{q}^2}{\alpha^2\lambda^2 kT}\right), \end{aligned} \tag{54}$$

where we have used (32) and have introduced E_0 from (17).

The time taken to pass through the critical range $(q_0 - dq_0, q_0)$ to dissociation at q_0 is $dq_0/\dot{q}$, if $\dot{q}$ is positive. Hence the proportion of molecules dissociating per second—the velocity constant—is

$$K = \int_0^\infty M(q_0, \dot{q})\, \dot{q}\, d\dot{q} = \frac{1}{2} \int_{-\infty}^{\infty} M(q_0, \dot{q})\, |\,\dot{q}\,|\, d\dot{q}. \tag{55}$$

(The second form follows as M is an even function of $\dot{q}$.) Using (54) we obtain

$$K = \nu \exp(-E_0/kT) \quad (\nu = \lambda/2\pi), \tag{56}$$

which agrees with the earlier result (34). This is to be expected, as (55), (53) are seen to be equivalent to the earlier basic formula (27), by reference to the form (50) of L_1. The interest of the present method lies in the use of the distribution $m(q_0, \dot{q})$ instead of using an explicit frequency L_1.

A 'transition state' formulation

In transition state theory a dissociation rate is formulated as the product of the equilibrium population of 'activated complexes' (molecules in the 'transition state' bordering on dissociation) and of the average rate of transition. The method just given can be translated into this form. Our transition state comprises all configurations with q increasing and in a short range $(q_0 - dq_0, q_0)$. The chance that a molecule with energies $(\epsilon_1, \ldots, \epsilon_n)$ is in this state may be represented by $\frac{1}{2}m(q_0)\,dq_0$, (40). Hence out of an equilibrium population (26), the concentration of activated complexes is $N^\ddagger$, where

$$N^\ddagger/N = \tfrac{1}{2} dq_0 \int_0^\infty \ldots \int_0^\infty m(q_0) f(\epsilon)\, d\epsilon_1 \ldots d\epsilon_n. \tag{57}$$

(The factor $\frac{1}{2}$ represents the choice of molecules with q increasing. Eyring and others transfer this factor to the mean velocity of transition.) Using (26) and (43), with (32) in the evaluation, we find

$$\frac{N^{\ddagger}}{N} = \frac{1}{2\alpha}\frac{dq_0}{\sqrt{(\pi kT)}}\exp(-E_0/kT). \tag{58}$$

Now the average velocity v in co-ordinate q through the transition state is, by reference to the definition of $M(q_0, \dot{q})$ above equation (53),

$$v = \int_{-\infty}^{\infty} M(q_0, \dot{q})\,|\,\dot{q}\,|\,d\dot{q} \bigg/ \int_{-\infty}^{\infty} M(q_0, \dot{q})\,d\dot{q}. \tag{59}$$

Hence using (54) we find $\quad v = \alpha\lambda\sqrt{(kT/\pi)}. \qquad (60)$

The average frequency of transition is v/dq_0: so the velocity constant is

$$K = \frac{N^{\ddagger}}{N}\frac{v}{dq_0} = \frac{\lambda}{2\pi}\exp(-E_0/kT), \tag{61}$$

as before. The equivalence of this method with the first approach of this section is seen by reference to (49). The 'transition state' formulation appears less direct because of the separate averaging of v in (59).

7. Generalization of Pelzer's calculation

Pelzer (1933) calculated the unimolecular rate for a chain molecule, with the potential energy of the form $V = \Sigma b_s q_s^2$, by a direct statistical method not involving normal vibrations. We shall extend his method to the general energies of (5). Transition state formalism will be used, as this does not here involve any complication.

In a population of N molecules in statistical equilibrium, the number with q_s and $\dot{q}_s$ in specified ranges dq_s, $d\dot{q}_s$ ($s = 1, \ldots, n$) at any time is assumed to be $N\,dg$, with

$$dg = C^{-1}\exp\{-(T'+V)/kT\}\,dq_1\ldots d\dot{q}_n, \tag{62}$$

where† $\quad C = \int\ldots\int \exp\{-(T'+V)/kT\}\,dq_1\ldots d\dot{q}_n, \qquad (63)$

and the energies T' and V are given by (5). The fraction of molecules 'activated', that is with $\dot{q}_1$ positive and q_1 in the transition region $(q_0 - dq_0, q_0)$, is

$$\frac{N^{\ddagger}}{N} = \tfrac{1}{2}dq_0\frac{\int\ldots\int \exp(-V_0/kT)\,dq_2\ldots dq_n}{\int\ldots\int \exp(-V/kT)\,dq_1\ldots dq_n}, \tag{64}$$

where V_0 denotes V with $q_1 = q_0$. The mean co-ordinate velocity, $v = \overline{|\dot{q}_1|}$, at the transition state is

$$v = \frac{\int\ldots\int \exp(-T'/kT)\,|\,\dot{q}_1\,|\,d\dot{q}_1\ldots d\dot{q}_n}{\int\ldots\int \exp(-T'/kT)\,d\dot{q}_1\ldots d\dot{q}_n}. \tag{65}$$

---

† The limits $(-\infty, +\infty)$ are to be understood in all integrals in §§ 7, 8, 9 where no limits are specified.

The unimolecular velocity constant is, as in § 6, the product of $N^{\ddagger}/N$ and the mean transition frequency:

$$K = \frac{N^{\ddagger}}{N}\frac{v}{dq_0}. \tag{66}$$

This could be written by (64), (65), or derived directly in imitation of Pelzer, as

$$K = \tfrac{1}{2}C^{-1}\int\ldots\int \exp\{-(T'+V_0)/kT\}\,|\,\dot{q}_1\,|\,dq_2\ldots dq_n\,d\dot{q}_1\ldots d\dot{q}_n. \tag{67}$$

We proceed to evaluate the factors (64), (65) of (66). In (64) we write

$$q_1 = x_1, \quad q_s = x_s + B_{1s}x_1/B_{11} \quad (s = 2, \ldots, n), \tag{68}$$

where B_{rs} is the co-factor of b_{rs} in $B \equiv \|\,b_{rs}\,\|$. This gives in (5)

$$2V = Bx_1^2/B_{11} + \sum_2^n\sum b_{rs}x_rx_s,$$

and in particular, when $q_1 = q_0$, the potential energy is, by (18),

$$V_0 = E_0 + \tfrac{1}{2}\sum_2^n\sum b_{rs}x_rx_s. \tag{69}$$

This identifies E_0 again as the minimum energy at the transition state, since the sum in (69) is essentially positive. By (68), (64) becomes

$$\begin{aligned} N^{\ddagger}/N &= \tfrac{1}{2}dq_0\exp(-E_0/kT)\Big/\int\exp(-Bx_1^2/2B_{11}kT)\,dx_1 \\ &= \tfrac{1}{2}dq_0\exp(-E_0/kT)\sqrt{(B/2\pi B_{11}kT)}. \end{aligned} \tag{70}$$

This agrees, by (11), with the result (58) of § 6.

To evaluate (65) we use the similar transformation

$$\dot{q}_1 = v_1, \quad \dot{q}_s = v_s + A_{1s}v_1/A_{11} \quad (s = 2, \ldots, n),$$

where A_{rs} is the co-factor of a_{rs} in $A \equiv \|\,a_{rs}\,\|$. This gives in (5)

$$2T' = Av_1^2/A_{11} + \sum_2^n\sum a_{rs}v_rv_s.$$

Thus (65) becomes

$$\begin{aligned} v &= \int|\,v_1\,|\,e^{-\phi}\,dv_1\Big/\int e^{-\phi}\,dv_1 \quad (\phi \equiv Av_1^2/2A_{11}kT) \\ &= \sqrt{(2kTA_{11}/\pi A)}, \end{aligned} \tag{71}$$

which agrees with (60) of § 6 by virtue of (11).

The results (70), (71) for the factors of K give for (66)

$$K = \nu\exp(-E_0/kT), \quad \nu = \frac{1}{2\pi}\left(\frac{A_{11}B}{AB_{11}}\right)^{\frac{1}{2}}, \tag{72}$$

which by (12) is the same as the results (34) of § 3 and (61) of § 6. Here, however, ν has not arisen as an 'average frequency', but as the combination (72) of determinants. The form (72) leads naturally to the 'particular case' discussed in § 2, and shows by reference to (22) that the unimolecular reaction rate for the breaking of an isolated bond between two atoms of a large molecule is the same as the rate for an imaginary diatomic molecule of these atoms with the same bond-strength (Pelzer 1947; Slater 1947*b*).

8. RELATIONS OF THE METHODS

The parallelism of the transition state methods in § 7 and at the end of § 6 was indicated by the identification of the corresponding expressions for $N^\ddagger/N$ (70) and v (71). The method of § 6 is a bridge between the sharply contrasted methods of §§ 3 and 7. It is of interest to complete the cycle by transforming the basic formula (67) of § 7 into that (27) of § 3.

Since in (67) V_0 denotes V with $q_1 = q_0$,

$$\exp(-V_0/kT) = \frac{1}{2\pi}\iint \exp\left[-\frac{V}{kT}+i(q_1-q_0)\xi\right]dq_1 d\xi.$$

This gives for (67)

$$K = \tfrac{1}{2}(2\pi C)^{-1}\int\ldots\int \exp\{-(T'+V)/kT\}\exp\{i(q_1-q_0)\xi\}\,|\dot{q}_1|\,dq_1\ldots d\dot{q}_n d\xi. \tag{73}$$

We transform from $(q_r, \dot{q}_r)$ to normal mode energies ϵ_s and phase angles θ_s, where (compare (8))

$$q_r = \Sigma\alpha_{rs}\sqrt{\epsilon_s}\cos 2\pi\theta_s, \quad \dot{q}_r = -\Sigma\alpha_{rs}\lambda_s\sqrt{\epsilon_s}\sin 2\pi\theta_s. \tag{74}$$

As this transformation may be compounded of linear transformations together with the contact transformation from normal co-ordinates and momenta to action and angle variables, it has a constant Jacobian. Applying (74) to the integral in (73) and also to the integral (63) for C, we obtain

$$K = \tfrac{1}{2}(2\pi C')^{-1}\int d\xi\int_0^\infty d\epsilon_1\ldots\int_0^1 d\theta_1\ldots\exp\{i(F-q_0)\xi\}\,|F'|\,\exp(-\Sigma\epsilon_s/kT), \tag{75}$$

where F, F' denote (39), (44) with $a_s = \alpha_s\sqrt{\epsilon_s}$ and

$$C' = \int_0^\infty d\epsilon_1\ldots\int_0^1 d\theta_1\ldots\exp(-\Sigma\epsilon_s/kT) = (kT)^n.$$

Hence by the formula (52) for the frequency $2L_1$,

$$K = (kT)^{-n}\int_0^\infty\ldots\int_0^\infty L_1\exp(-\Sigma\epsilon_s/kT)\,d\epsilon_1\ldots d\epsilon_n, \tag{76}$$

which by (26) is the basic formula (27) of § 3.

This identification raises the question of the status, in the formulation § 7, of the linear independence of the frequencies, imposed in § 3. It was suggested in § 4 that (*a*) an energized molecule starts with phases concentrated towards $\frac{1}{4}$ or $\frac{3}{4}$, and that (*b*) if the frequencies are linearly independent this distribution will be smoothed out during the energized lifetime. The corresponding picture in the $(q, \dot{q})$ or (q, p) phase space of § 7 (with $p_r = \partial T'/\partial\dot{q}_r$, the generalized momentum) is that (*a*) an energized molecule starts with comparatively small q's and large p's, and that (*b*) these values spread out over available phase-space. It is known that if the frequencies considered as the 'average motions' of the system (compare Epstein 1936) are linearly independent, then any trajectory fills the ergodically available phase space uniformly. This corresponds to the distribution (62) of § 7, extended as in (63) over all phase space: this distribution would therefore not be justified (with our initial concentration of q, p) if the ν_s were not linearly independent. There is thus a close correspondence between the role of linear independence in §§ 3 and 7. The transformation (74) preserved equality of weighting of elements of phase space, as is required if the correspondence is to hold.

9. Transition velocity and partition functions

The present model will now be used to illustrate some concepts of transition state theory, in particular the transition velocity and the partition function of the activated complex. This theory has been fully expounded by Glasstone, Laidler & Eyring (1941), to whom we refer as 'G.L.E.'.

In our model the transition velocity v, defined by (59) or (65), takes the form (71). This corresponds to the formula of G.L.E. (p. 187, first line)

$$\bar{\dot{x}} = \tfrac{1}{2}\sqrt{(2kT/\pi m^*)}, \tag{77}$$

if we identify their m^* with A/A_{11} in (71). (The extra factor $\frac{1}{2}$ in (77), absent from (71), was included in the activated population (57), as we noted in § 6.) G.L.E. call m^* in (77) the 'effective mass of the activated complex in the reaction co-ordinate'. Now if the kinetic energy T' of (5) is expressed in terms of the momenta $p_r = \partial T'/\partial \dot{q}_r$, the term in p_1^2 is $\frac{1}{2}p_1^2 A_{11}/A$. Thus A/A_{11} is a mass-coefficient associated with the breaking co-ordinate q_1. In the important case in § 2 where q_1 is the bond extension between two atoms m_1, m_2, then by (19) A/A_{11} is the 'reduced mass' $m_1 m_2/(m_1+m_2)$. This shows that in this case the effective mass of the activated complex depends only on the adjacent atoms and not on the rest of the molecule.

We turn now to partition functions, and imitate the notation of G.L.E. (pp. 188–9) to facilitate comparison. The relative concentration $N^\ddagger/N$ ((57) or (64)) of 'activated complexes' equals the ratio $F'_\ddagger/F$ of the partition functions of activated and normal molecules, so that the rate (66) may be written

$$K = \frac{v}{dq_0}\frac{F'_\ddagger}{F}. \tag{78}$$

The classical partition function F for the energies (5) is

$$F = h^{-n}\int\cdots\int \exp\{-(T'+V)/kT\}\,dp_1\ldots dq_n = \prod_1^n (kT/h\nu_s), \tag{79}$$

where Planck's constant h is introduced to equate classical to quantal weighting of phase space, and the ν_s are the normal frequencies. The partition function $F'_\ddagger$ for the activated complex (q_1 increasing and in (q_0-dq_0, q_0)) is

$$F'_\ddagger = \tfrac{1}{2}h^{-n}dq_0\int\cdots\int \exp\{-(T'+V_0)/kT\}\,dp_1\ldots dp_n\,dq_2\ldots dq_n,$$

where V_0 denotes V with $q_1 = q_0$, as in § 7. By using (68) we find

$$F'_\ddagger = \tfrac{1}{2}h^{-n}dq_0(2\pi kT)^{n-\frac{1}{2}}\sqrt{(A/B_{11})}\exp(-E_0/kT). \tag{80}$$

This may be arranged in the form

$$F'_{\ddagger} = f^{\ddagger}_{tr(1)} F_{\ddagger} \exp(-E_0/kT), \tag{81}$$

where

$$f^{\ddagger}_{tr(1)} = \frac{dq_0}{2h}\sqrt{(2\pi kTA/A_{11})}, \tag{82}$$

$$F_{\ddagger} = (2\pi kT/h)^{n-1}\sqrt{(A_{11}/B_{11})} = \prod_{2}^{n} (kT/h\nu'_s), \tag{83}$$

by the use in (83) of the formula (14) for the frequencies $\nu'_s = \lambda'_s/2\pi$ of the system with q_1 fixed. Now by (71), (82),

$$\frac{v}{dq_0} f^{\ddagger}_{tr(1)} = \frac{kT}{h}. \tag{84}$$

Thus the rate (78) becomes, by (81), (84),

$$K = \frac{kT}{h}\frac{F_{\ddagger}}{F}\exp(-E_0/kT), \tag{85}$$

which may also be written by (79), (83)

$$K = \frac{\nu_1\nu_2\ldots\nu_n}{\nu'_2\ldots\nu'_n}\exp(-E_0/kT). \tag{86}$$

This repeats, by (15), our earlier result $\nu\exp(-E_0/kT)$; it also corresponds explicitly to the classical approximation of G.L.E. (p. 193) if only vibrational motion is considered.

We shall examine more closely the factors of the previous form (85), which is a strict imitation of the general rate formula of G.L.E. (p. 189). The first factor kT/h, called by them a 'universal constant' of rate processes at a given temperature, was constructed in (84) out of v/dq_0, the mean transmission frequency, and $f^{\ddagger}_{tr(1)}$, which by (82) has the form of a translational partition function for motion in the reaction co-ordinate with A/A_{11} as the 'effective mass'. On our model kT/h appears to be at most a formal construct, just as the factor h, as noted under (79), plays only a formal role in the classical theory and does not affect the total rate $\nu\exp(-E_0/kT)$.

Of the remaining factors of (85), $F_{\ddagger}\exp(-E_0/kT)$ is the partition function for the molecule with q_1 fixed at q_0. For by (68), (69) the energies are then

$$V = E_0 + \tfrac{1}{2}\sum_{2}^{n}\sum b_{rs}(q_r - q_{r0})(q_s - q_{s0}), \quad T' = \tfrac{1}{2}\sum_{2}^{n}\sum a_{rs}\dot{q}_r\dot{q}_s, \tag{87}$$

where $q_{r0} = B_{1r}q_0/B_{11}$ gives the configuration of least energy (E_0) at the transition state. As the frequencies of the system (87) are the ν'_s defined under equation (15), $F_{\ddagger}$ (83) is the corresponding vibrational partition function, and $\exp(-E_0/kT)$ represents the difference in zero energy level between this system and the normal molecule.

This completes the parallel with G.L.E.'s treatment, showing the rate as a product (85) with factors kT/h and a modified partition function representing vibrations at the transition point. This analysis has an artificial character here, although formulae of the type (85) have been much used in the development of the general theory.

10. POTENTIAL ENERGY DIAGRAMS

The dissociation model is illustrated here by the potential energy diagrams used in transition state theory. This leads to a discussion of the earlier assumption that the vibrations are strictly harmonic.

Graphical representations are more simply interpreted if we choose co-ordinates q'_s such that the kinetic energy is a sum of squares (G.L.E. chapter III). To preserve the identity of the breaking co-ordinate q_1, we choose a transformation (with unit Jacobian) which gives in (5)

$$T' = \frac{1}{2}\frac{A}{A_{11}}\sum_1^n \dot{q}'^2_s \quad \text{with} \quad q'_1 \equiv q_1. \tag{88}$$

This leaves the potential (5) in a general form

$$V = \tfrac{1}{2}\sum_1^n\sum b'_{rs}q'_r q'_s. \tag{89}$$

The internal motion of the molecule can now be represented by the motion of a particle P of mass A/A_{11} in an n-dimensional space with rectangular co-ordinates $(q'_1, \ldots, q'_n)$ under the potential (89).

The equipotential curves are illustrated in figure 1 for the case $n = 2$. The curves are similar ellipses, centred at the origin (the equilibrium configuration), and increasing in size with V. The principal axes represent the paths of P in one or other of the normal modes of vibration. If the system has energies ϵ_1, ϵ_2 in the two normal modes, with for example $\epsilon_1 + \epsilon_2 = 10$, then P describes a Lissajous curve bounded by a rectangle (indicated in the figure by a broken line) inscribed in the ellipse $V = 10$. If the frequencies ν_1, ν_2 are linearly independent, the Lissajous curve is not closed, but covers in time the whole rectangle: this is the fundamental role of the condition of linear independence. Different values of ϵ_1, ϵ_2 with the same sum 10 give motions in other rectangles inscribed in the same ellipse. The 'transition state' is a narrow strip $q_1 = (q_0 - dq_0, q_0)$, and the molecule dissociates when P crosses the line $q_1 = q_0$ to the right. The 'activation energy' in figure 1 is $E_0 = 20$, since the ellipse $V = 20$ touches this line. The point A of contact is the transition point of least energy: its co-ordinates were given under equation (87).

This picture may be extended mentally from the case $n = 2$ to the large values of n characterizing polyatomic molecules. If E_0 is large, most energetic molecules will have energies not much larger than E_0, so that their equipotentials will contain only a small region of the hyperplane $q_1 = q_0$. Thus the Lissajous curves may perform many loops before crossing this plane. This illustrates the result of § 4 that the dissociation frequency L is small even for a favourable distribution of the typical high energy $E_0 + kT$, when n is large.

Returning to the illustrative case $n = 2$, we discuss finally the harmonicity of the vibrations. If the molecule is to dissociate at $q_1 = q_0$, V must diminish as q_1 increases beyond q_0, so that the elliptic contours cannot continue into this region. If, moreover, the contours are elliptic right up to $q_1 = q_0$, then $\partial V/\partial q_1$ is strictly positive to the left and is negative to the right of this line. Our model thus implies a discontinuity in the gradient of V. This may be a useful approximation, but for continuity we require $\partial V/\partial q_1 = 0$ on $q_1 = q_0$. Thus the elliptic contours, and the potential function (5) or (89), must be modified for $q_1 < q_0$; the complete equipotentials must therefore be of the well-known type sketched in figure 2, showing the characteristic 'col' at the activation point A.

This 'distortion' of the contours for q_1 less than q_0 implies that the vibrations cannot remain harmonic up to the point of rupture. The error of our harmonic model will, however, be small if, for example, V is modified only very near $q_1 = q_0$, with a rapid change of gradient. In a one-dimensional diffusion model Kramers (1940) has compared passage over a rounded barrier with passage over a sharp peak of V. The present model might be modified by changing the form of V to round off its values near $q_1 = q_0$; this modification could for example be introduced into the formula (67) for the dissociation rate. The general effect of such a modification has been discussed by Evans & Rushbrooke (1945) with reference to entropy changes.

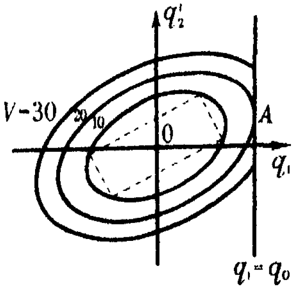


FIGURE 1. Elliptical equipotential curves.

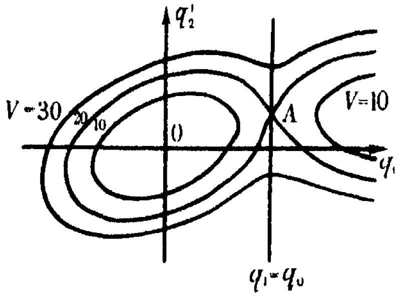


FIGURE 2. Modified equipotential curves.

11. SPECIFIC DISSOCIATION RATES

In conclusion a brief account is given of the analysis of the dissociation rate into the contributions of molecules of specified total energy, and the results are compared with Kassel's theory. The method of § 3 is used as a basis.

Molecules whose internal energy $\Sigma\epsilon_s$ lies in a short range $(E, E+dE)$ will be called 'E-molecules'. The number of these in an equilibrium distribution (26) is $Nf_E\,dE$, where

$$f_E\,dE = \int_{E<\Sigma\epsilon_s<E+dE}\cdots\int f(\epsilon)\,d\epsilon_1\ldots d\epsilon_n = \frac{E^{n-1}e^{-E/kT}\,dE}{\Gamma(n)\,(kT)^n}. \tag{90}$$

If $NK_E\,dE$ is the number of E-molecules dissociating per second, then the overall rate K, (27), is

$$K = \int K_E\,dE = \int\cdots\int Lf(\epsilon)\,d\epsilon_1\ldots d\epsilon_n.$$

Hence, as E represents $\Sigma\epsilon_s$, we have by (26)

$$\begin{aligned}\int_0^\infty K_E\,e^{-Ez}\,dE &= (kT)^{-n}\int_0^\infty\cdots\int_0^\infty L\exp\left\{-\left(z+\frac{1}{kT}\right)\Sigma\epsilon_s\right\}d\epsilon_1\ldots d\epsilon_n\\ &= \nu(1+zkT)^{-n}\exp\left\{-E_0\left(z+\frac{1}{kT}\right)\right\},\end{aligned} \tag{91}$$

where ν, E_0 are defined by (9), (17) and the integration has imitated the process of § 3. Denoting the last expression in (91) by $\phi(z)$, we have as the Laplace transform

$$K_E = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\phi(z)\,e^{Ez}\,dz \quad (c \geqq 0).$$

Evaluation gives

$$\left.\begin{aligned} K_E &= \frac{\nu e^{-E/kT}(E-E_0)^{n-1}}{\Gamma(n)\,(kT)^n} &&(E > E_0),\\ &= 0 &&(E < E_0).\end{aligned}\right\} \tag{92}$$

The specific dissociation rate, or fraction of E-molecules dissociating per second, is by (90)

$$K'_E = \frac{K_E}{f_E} = \nu\left(\frac{E-E_0}{E}\right)^{n-1} \quad (E > E_0). \tag{93}$$

This represents the average of L (§ 3) over all distributions $(\epsilon_1, \ldots, \epsilon_n)$ of a total energy E, including those distributions for which (25) is not satisfied and L vanishes. A similar application of Laplace transforms to the results of §§ 6 or 7 gives the number of E-molecules with specified values of q, or of q and $\dot{q}$. The average velocity of E-molecules through the transition state is found, for example, to be

$$v_E = \frac{\Gamma(n-\frac{1}{2})}{\Gamma(n)}\left(\frac{2(E-E_0)}{\pi A/A_{11}}\right)^{\frac{1}{2}}. \tag{94}$$

The specific rate (93) is noteworthy for its formal identity with the corresponding classical rate of Kassel (1932). He obtains the factor $[(E-E_0)/E]^{n-1}$ as the chance that, of a total energy E distributed between n equivalent oscillators, more than E_0 is in one (breaking) oscillator. Instead of the precise factor ν in (93), he has an arbitrary constant representing the frequency of interchange of energy between oscillators. The resemblance of this model to the present model of dissociation is thus not close, although the common formula (93) ensures the same form, $\nu\exp(-E_0/kT)$, of high-pressure rate.

The formula (93) could be used to extend the theory to low-pressure dissociation, in exact imitation of Kassel's method. On our model, however, this use of K'_E would not be accurate. At low pressures the concentration of energetic molecules is depleted by dissociation, and for a given total energy E those distributions $(\epsilon_1, \ldots, \epsilon_n)$ giving the largest values of L are the most depleted, so that the use of an average K'_E of L is incorrect. The low-pressure rate can, however, be formulated precisely as an integral involving L as a function of the energies $\epsilon_1, \ldots, \epsilon_n$. Kac's formula (31) for L is unsuitable for integration in this case, but the rate can be estimated by using my previous approximate formula (Slater 1939, equation (8)) for L.

I am indebted to Professor M. G. Evans and Dr G. S. Rushbrooke, who suggested the translation of the theory into 'transition state' form, and to Mr H. D. Ursell, for many helpful discussions. I am grateful also to Professor P. Stein for communicating to me some unpublished theorems on trigonometrical sums.

REFERENCES

Epstein, P. S. 1936 Article U (vol. II) in *Commentary on ... J. Willard Gibbs.* Yale University Press.
Evans, M. G. & Rushbrooke, G. S. 1945 *Trans. Faraday Soc.* **41**, 621.
Glasstone, S., Laidler, K. J. & Eyring, H. 1941 *The theory of rate processes.* New York: McGraw-Hill Book Co.
Kac, M. 1943 *Amer. J. Math.* **65**, 609.
Kassel, L. S. 1932 *Kinetics of homogeneous gas reactions*, chapter v. New York: Chemical Catalog Co.
Kramers, H. A. 1940 *Physica*, **7**, 284.
Pelzer, H. 1933 *Z. Elektrochem.* **39**, 608.
Pelzer, H. 1947 *Nature*, **160**, 576.
Polanyi, M. & Wigner, E. 1928 *Z. phys. Chem.* A, **139**, 439.
Rice, S. O. 1945 *Bell Syst. Tech. J.* **24**, 46.
Slater, N. B. 1939 *Proc. Camb. Phil. Soc.* **35**, 56.
Slater, N. B. 1947*a* *Nature*, **159**, 264.
Slater, N. B. 1947*b* *Nature*, **160**, 576.
Watson, G. N. 1922 *Theory of Bessel functions.* Cambridge University Press.
Wintner, A. 1932 *Math. Z.* **36**, 618.

The transmutation of magnesium, aluminium and silicon by deuterons

By H. R. ALLAN, *Trinity College, University of Cambridge*
AND CHRISTIANE A. WILKINSON, *Newnham College, University of Cambridge*

(*Communicated by J. D. Cockcroft, F.R.S.—Received* 31 *December* 1947)

A study has been made of the protons emitted by magnesium, aluminium and silicon targets when bombarded by deuterons of energy between 700 and 1000 keV. The Q values of these (d, p) reactions have been accurately determined. Observations of the induced radioactivity in the targets and comparisons with the Q values in analogous nuclei permit tentative assignments to the isotopes responsible. The Q values and assignments are:

$$\begin{array}{ll}
{}^{24}\text{Mg}\ (d, p)\ {}^{25}\text{Mg} & Q = 5{\cdot}03\ \text{MeV}\\
{}^{25}\text{Mg}\ (d, p)\ {}^{26}\text{Mg}^{*} & Q = 4{\cdot}45\ \text{MeV (to excited state)}\\
{}^{26}\text{Mg}\ (d, p)\ {}^{27}\text{Mg} & Q = 4{\cdot}05\ \text{MeV}\\
{}^{27}\text{Al}\ (d, p)\ {}^{28}\text{Al} & Q = 5{\cdot}46\ \text{MeV}\\
 & \left.\begin{array}{l} = 4{\cdot}46\ \text{MeV}\\ = 3{\cdot}98\ \text{MeV}\\ = 3{\cdot}31\ \text{MeV}\end{array}\right\}\ \text{(to excited states)}\\
{}^{28}\text{Si}\ (d, p){}^{29}\text{Si} & Q = 6{\cdot}16\ \text{MeV}\\
 & \left.\begin{array}{l} = 4{\cdot}87\ \text{MeV}\\ = 3{\cdot}75\ \text{MeV}\end{array}\right\}\ \text{(to excited states)}\\
{}^{30}\text{Si}\ (d, p){}^{31}\text{Si} & Q = 4{\cdot}16\ \text{MeV}
\end{array}$$

The yield of the proton group attributed to ^{25}Mg (d, p) ^{26}Mg* shows a resonance at 955 kV bombarding voltage, as does also the yield of radioactive ^{26}Al formed in the reaction ^{25}Mg (d, n) ^{26}Al. This indicates a resonance level in the compound nucleus ^{27}Al at an excitation energy of 17 MeV. The maximum energy of the β^{+} rays from ^{26}Al was measured as 2·8 MeV.

The Q values listed above lead to a set of nuclear mass values, but these mass values are not always in satisfactory agreement with the published Q values for certain other reactions.

1. INTRODUCTION

The nuclear masses of elements in the Periodic Table between neon and sulphur are not known with the same accuracy as those of lighter elements. The masses of the nuclei ^{20}Ne, ^{27}Al and ^{32}S have been determined with an error of less than 10^{-4} mass unit with the mass spectrograph, but mass values for many other nuclei in this region depend to a large extent on disintegration data (Livingston & Bethe 1937; Pollard 1940; Fluegge & Mattauch 1943). The accuracy of these published values depends directly upon the precision of the measurement of the energy releases (Q values) in (α, p) and (d, p) reactions, which depends in turn upon the accuracy with which the energy of the bombarding particle can be defined and that of the emitted proton determined. For (α, p) reactions, the published Q values have large probable errors mainly because natural α-particle sources have low intensity, but in (d, p) reactions considerably greater accuracy can be obtained.

This paper presents the results of measurements of the Q values of the (d, p) reactions in magnesium, aluminium and silicon; the errors are estimated to be about 0·05 MeV (5×10^{-5} mass unit). A short account of some of this work has been published previously (Allan & Clavier 1946). The Q values themselves are all of the order of 5 MeV, and the experiments therefore provide also a means of investigating the energy levels of the residual heavy nuclei up to this level of excitation. In the past this type of work has been confined chiefly to elements of lower atomic number, for which the level spacing is about 1 MeV; with increasing mass the spacing is expected to decrease.

The (d, p) disintegrations of elements between neon and sulphur have hitherto been investigated almost exclusively by using a cyclotron as the source of bombarding particles (McMillan & Lawrence 1935; Schultz, Davidson & Ott 1940; Pollard 1940; Pollard & Humphreys 1941). In the experiments of Pollard and his collaborators the energy of the bombarding particles was between 3 and 4 MeV, and could be defined to an accuracy of 0·18 MeV (Martin 1947). This definition can be improved by the use of a high voltage equipment for accelerating the deuterons (Cockcroft & Walton 1934; Burcham & Smith 1938; Murrell & Smith 1939), but only energies of up to 1 or 2 MeV may then be available which means that the incident deuteron must penetrate through the potential barrier of the bombarded nucleus. For the elements under investigation the probability of barrier penetration by 1 MeV deuterons is about 0·001, and large bombarding currents are required to give a sufficient yield; in the present work, currents of about 50 μA at an accelerating voltage of 930 kV were used. Under these conditions, lighter elements present on the targets as contaminants give a relatively copious yield of neutrons, which are particularly troublesome in that they produce a high background counting rate in the chambers used for detecting the protons; statistical fluctuations in this background set a lower limit to the detectable yield of protons. Elements of atomic number greater than that of phosphorus give too small a yield and cannot easily be investigated at a voltage less than 1000 kV.

In § 2 (a) of this paper the experimental methods used for measuring proton ranges are described; in § 2 (b) the results of these measurements are given, and the

Q values to be associated with the (d,p) reactions in magnesium, aluminium and silicon are tabulated, while the excitation functions are considered in § 2 (*c*). § 3 deals with the radioactivity of the targets after bombardment. In § 4 a general discussion on the assignment of the observed Q values is given, and a tentative table of nuclear masses put forward.

2. Proton range measurements

(*a*) *Method*

The experiments were carried out with the Cambridge 1 million volt high-tension apparatus; the magnetically resolved deuteron beam of $50\,\mu$A at 930 ± 15 kV fell on a target area of about 1 cm.^{2}. The targets to be bombarded were prepared on gold-plated copper buttons, which screwed into the water-cooled copper block B in the target chamber (figure 1). This chamber was substantially the same as that used by Burcham & Smith (1938). Rotation of the vacuum-tight ground joint G enabled any one of three target buttons to be brought into the path of the ion beam.

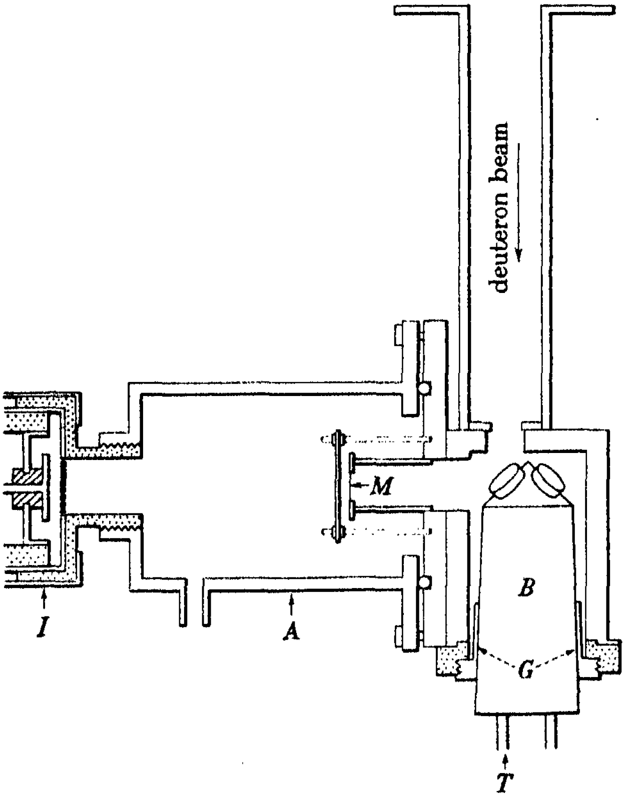


Figure 1. Apparatus for measuring proton ranges.

░, ebonite; ▨, polythene; I, ionization chamber; A, air cell; T, target chamber.

Magnesium and aluminium were deposited on the target buttons by evaporation of the metal *in vacuo*, and silicon was cut in the form of a slab from a fused lump of the element.

Particles emitted from the target at right angles to the deuteron beam emerged from the vacuum system through two mica windows in the target chamber walls (*M*, and another out of the plane of the diagram in figure 1), and then entered the detecting chambers. In the first experiments with magnesium and aluminium these were two shallow air-filled ionization chambers, 3·5 mm. deep, each connected to a linear amplifier, discriminator and scaling unit. The voltage pulses from the amplifier due to protons entering the ionization chamber were greater than the noise level of the amplifier only when the proton residual range was less than 2 cm., the maximum amplitude being about twice the noise level. The discriminator was adjusted so that the amplifier noise was just not recorded; thus only protons entering the chamber within 2 cm. of the end of their range were counted and a numbers-range curve, obtained by varying the absorption between target and ionization chamber, showed several peaks corresponding to the various proton groups. To allow for fluctuations in the counting rate due to changes in the deuteron beam current, and for changes in the target itself under bombardment, one of the ionization chambers was used as a monitor (Burcham & Smith 1938). The absorption between the target and the monitor was such that protons belonging to one of the groups emitted by the target element were counted. Readings with the other (search) ionization chamber were corrected to a standard level by using the monitor counts.

The search ionization chamber was connected to the target chamber by a cell containing dried air, and the absorption between it and the target could be varied by placing mica sheets in front of the target chamber window and by adjusting the air pressure. The temperature remained between 21 and 22° C throughout the experiments. A pressure of 1 cm. of mercury in the cell represented an effective stopping power between target and ionization chamber of 0·101 ± 0·001 cm. air at 15° C and 760 mm. The stopping powers of the mica absorbers were found both by weighing and by noting the changes of air pressure in the air cell equivalent to the addition of absorbers, the numbers-range curve for a well-known group of protons being used in this comparison. The use of an air cell is preferable to movement of the ionization chamber as a whole, since it ensures a definite and reproducible geometrical arrangement between the target and the ionization chamber.

Figure 2 shows the numbers-range curve obtained for the proton group resulting from the bombardment of beryllium by deuterons. From this curve an extrapolated range (28·49 ± 0·3 cm. of air) was obtained in the usual way by producing the straight descending portion of the curve to intersect the background counting level. This extrapolated range was corrected by the method of Livingston & Bethe (1937) for good geometry in order to obtain the mean range of the proton group (28·51 ± 0·4 cm. of air). The method used includes a correction of − 0·28 cm. of air for range straggling, angle straggling, and thickness of target, and a correction of + 0·30 cm. of air for variation of the stopping power of mica with velocity. The energy of the proton group was obtained from the mean range by using the range-energy relation given by Livingston & Bethe. The Q value for the beryllium (d,p) reaction was found by this method to be $4{\cdot}59 \pm 0{\cdot}05$ MeV in good agreement with previously accepted values (Oliphant, Kempton & Rutherford 1935; Livingston & Bethe 1937). This agreement indicated that the method used for range measurement was trustworthy, and it was applied in the analysis of the proton groups found in the transmutation of the other elements investigated in this work.

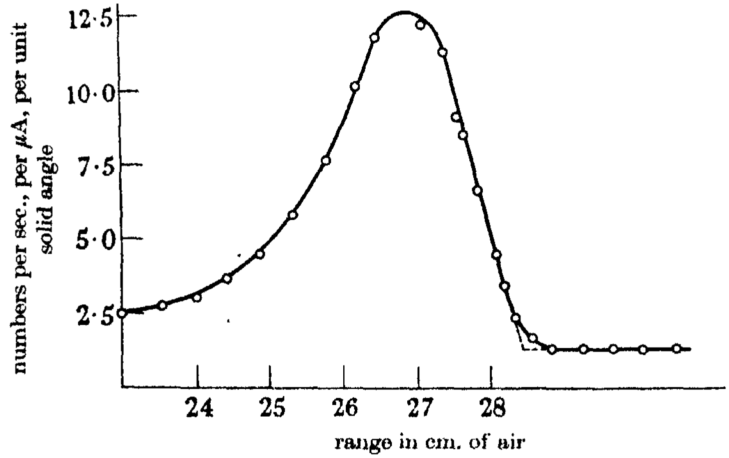


FIGURE 2. Protons from beryllium.

In the bombardment of magnesium and aluminium there was an undesirably high background counting rate in the ionization chambers due to neutron recoils. With the object of discriminating against such neutron recoils, the ionization chamber was replaced by a proportional counter and appropriate amplifiers, with which it was found possible to obtain pulses ten times the amplifier noise level. This permitted an increase in discriminator bias, to give both higher resolution in range for the proton groups and a lower background. The proportional counter was therefore used in the work on silicon where the proton yield is lower than for magnesium and aluminium; the target box of figure 1 was not used in this work, and the geometrical arrangement was not so good since the available intensity had to be increased as far as possible. Mica absorbers only were used to vary the absorption between the target and the proportional counter, and the ranges were determined by measuring the difference between the range of the group concerned and that of a well-known group. The performances of the ionization chamber and the proportional counter are compared in figure 3 (*b*), which gives the numbers-range curve for aluminium taken by both methods. The better resolution and lower background of the proportional counter are clearly shown.

(*b*) *Results*

The numbers-range curves for magnesium, aluminium and silicon bombarded by 930 keV deuterons are shown in figures 3 (*a*) and (*b*) and figure 4. Figures 3 (*c*) and (*d*) show similar curves for carbon and nitrogen, which are possible contaminants giving proton groups of comparable range. It is clear that the groups near 21 and 50 cm.

must probably be attributed to the well-known reactions $^{14}N\ (d, p)\ ^{15}N^*$, $^{13}C\ (d, p)\ ^{14}C$ respectively ($^{15}N^*$ indicating an excited state of the ^{15}N nucleus). Groups of range less than 16 cm. were not investigated owing to the presence of intense contamination groups from deuterium and oxygen.

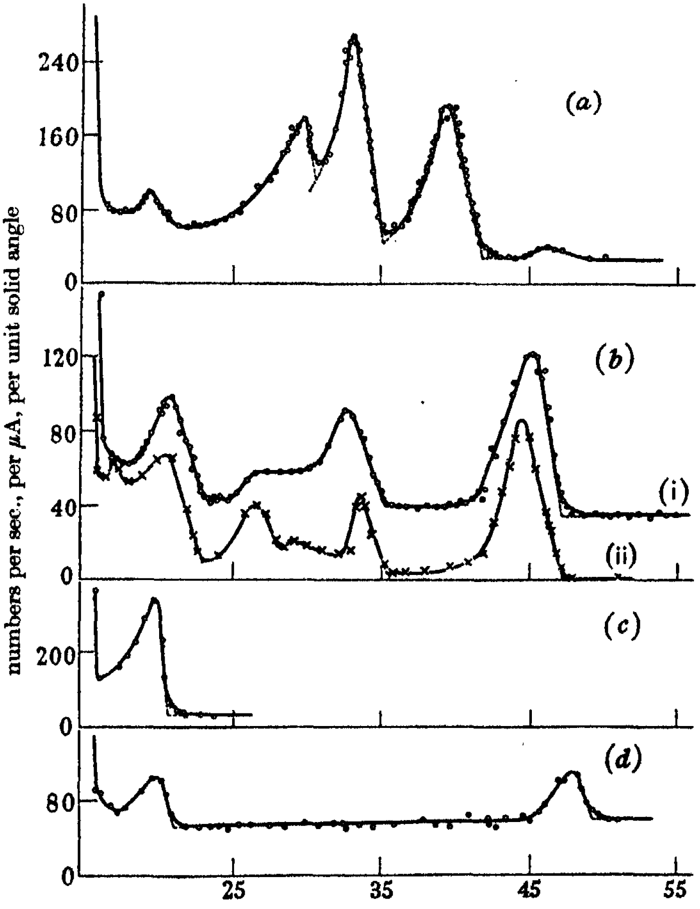


FIGURE 3. Proton groups from magnesium, aluminium, and contaminants. (*a*) Magnesium target; (*b*) aluminium target, (i) ionization chamber measurements, (ii) proportional counter measurements; (*c*) nitride target; (*d*) carbon target.

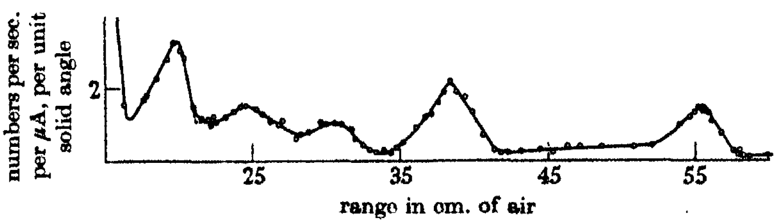


FIGURE 4. Proton groups from silicon.

Table 1 lists the mean ranges and Q values for the various groups (excluding those due to contaminants) in accordance with the assignment of the groups to the various isotopes as discussed in § 4.

TABLE 1

reaction	mean range (cm.)	Q value (MeV)	other workers' Q values (MeV)	reference
^{24}Mg (d, p) ^{25}Mg	42·2 ± 0·4	5·03 ± 0·05	5·3	Pollard; Pollard & Humphreys
^{25}Mg (d, p) ^{26}Mg*	35·4 ± 0·4	4·45 ± 0·05	8·15 6·3 5·15	
^{26}Mg (d, p) ^{27}Mg	30·85 ± 0·4	4·05 ± 0·05	—	—
^{27}Al (d, p) ^{28}Al	23·4 ± 0·4 29·3 ± 0·5 35·7 ± 0·4 48·0 ± 0·5	3·31 ± 0·05 3·98 ± 0·06 4·46 ± 0·05 5·46 ± 0·06	0·64 2·12 3·10 5·11 5·79 ± 0·3	McMillan & Lawrence
^{28}Si (d, p) ^{29}Si	27·8 ± 0·5 40·9 ± 0·5 57·9 ± 0·5	3·75 ± 0·06 4·87 ± 0·06 6·16 ± 0·06	— 5·7	— Pollard
^{29}Si (d, p) ^{30}Si	—	—	6·2 7·0 7·9 8·9 9·8	Pollard; Pollard & Humphreys
^{30}Si (d, p) ^{31}Si	32·4 ± 0·5	4·16 ± 0·06	—	—

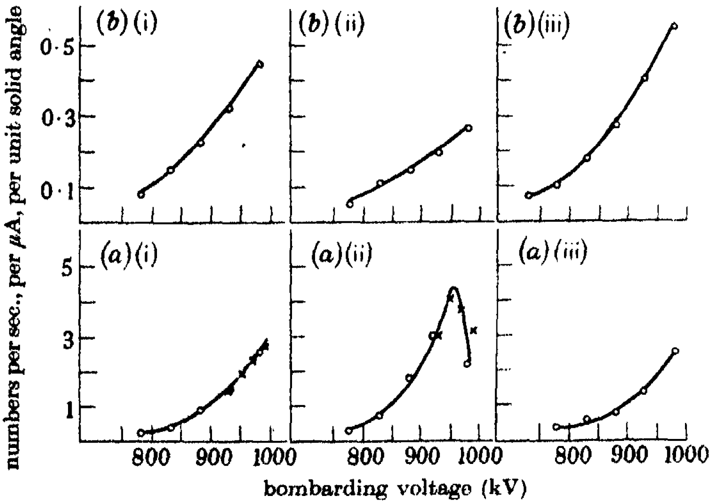


FIGURE 5. Excitation functions for proton groups. ○, ionization chamber Measurements; ×, proportional counter measurements; (*a*) magnesium, (*i*) range 31 cm., (*ii*) range 35 cm., (iii) range 42 cm.; (*b*) aluminium, (i) range 23 cm., (ii) range 36 cm., (iii) range 48 cm.

In the last column of table 1 the results of other workers are listed; some of these are calculated from the list of mass values published by Pollard (1940). Schultz, Davidson & Ott (1940), working with a cloud chamber and using 3·2 MeV deuterons, found indications of fine structure in the proton groups from aluminium, but this is not confirmed in the present experiments. The Q values calculated from the masses given by Pollard (1940) and by Fluegge & Mattauch (1943) for the magnesium and silicon isotopes suggest that groups of longer range (about 100 cm.) were missed in the present experiments, the reactions concerned being $^{25}\mathrm{Mg}\,(d,p)\,^{26}\mathrm{Mg}$ and $^{29}\mathrm{Si}\,(d,p)\,^{30}\mathrm{Si}$. A careful search was made in each bombardment for proton groups of ranges up to 150 cm., but none was found above the neutron background.

(c) *Excitation function*

The variation of the yield of the proton groups from magnesium and aluminium at bombarding voltages between 700 and 1000 kV was determined using thin targets (figure 5). The target thickness could not be estimated accurately but was probably less than 10 keV equivalent, as resonances observed in the reactions $^{25}\mathrm{Mg}\,(p,\gamma)\,^{26}\mathrm{Al}$ and $^{27}\mathrm{Al}\,(p,\gamma)\,^{28}\mathrm{Si}$ had widths of about 30 keV, which is of the same order as that of the beam energy distribution. The resonance apparent for one of the magnesium proton groups is referred to in § 4 (*b*).

3. RADIOACTIVITY

The deuteron bombardment of magnesium, aluminium and silicon may result in the formation of a number of different radioactive nuclei, as shown by the following reactions (Seaborg 1944):

(1)	$^{26}\mathrm{Mg}\,(d,p)\,^{27}\mathrm{Mg}$	10·2 min.	β^-, γ
(2)	$^{26}\mathrm{Mg}\,(d,\alpha)\,^{24}\mathrm{Na}$	14·8 hr.	β^-, γ
(3)	$^{25}\mathrm{Mg}\,(d,n)\,^{26}\mathrm{Al}$	7·0 sec.	β^+
(4)	$^{27}\mathrm{Al}\,(d,p)\,^{28}\mathrm{Al}$	2·4 min.	β^-, γ
(5)	$^{30}\mathrm{Si}\,(d,p)\,^{31}\mathrm{Si}$	170 min.	β^-

The nucleus ^{26}Al can also be formed from magnesium by proton bombardment in accordance with the reaction

(6) $^{25}\mathrm{Mg}\,(p,\gamma)\,^{26}\mathrm{Al}$.

If reactions (1) and (4) are responsible for certain groups of protons observed in § 2 (*b*), then the activity of the resultant nuclei, ^{27}Mg and ^{28}Al, should be detectable. The radioactivity of the targets after bombardment was therefore examined.

(a) *Magnesium*

After a day's bombardment, the magnesium target was removed from the target chamber and its β-activity followed for several days. The decay showed three distinct components of half-lives 10 min., 2·1 hr. and 14·8 hr. Part of the decay curve, with its 2·1 hr. and 14·8 hr. components, is shown in figure 6.

The 10 min. activity was found by a rough magnetic analysis to consist chiefly of positrons, presumably coming from ^{13}N formed by the reaction $^{12}\mathrm{C}\,(d,n)\,^{13}\mathrm{N}$ in the carbon contamination; about 1/20 of the activity, however, consisted of electrons, and was attributed to ^{27}Mg formed according to reaction (1).

To permit a comparison of the intensity of the ^{27}Mg activity with that of the observed proton groups the β-active target and an end-on β-counter were mounted in the same geometrical conditions as the target and ionization chamber during proton measurements. With allowances for decay and for the 10 min. positron activity, the number of β-particles that would have been counted at the time of ending the bombardment was found to be 7000 per min. This should also be the number of protons per min. from the reaction ^{26}Mg (d,p) ^{27}Mg; a direct determination for the 31 cm. group gave only 2000 per min., which is not unacceptable considering the rough nature of the comparison.

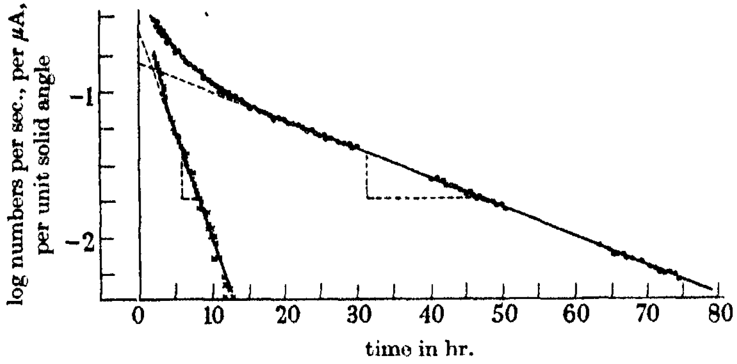


FIGURE 6. Decay curve for magnesium target after deuteron bombardment.

The 14·8 hr. activity was attributed to ^{24}Na, and the assignment confirmed by determining the energy of the β-particles from their range in aluminium. There is the possibility that the ^{24}Na might have been produced not by reaction (2) but by the alternative process ^{24}Mg (n,p) ^{24}Na from neutrons formed in the ^{2}H (d,n) ^{3}He reaction with adsorbed deuterium on the target. This possibility was excluded by mounting a second magnesium target close behind the first, where it could not be hit directly by the deuteron beam, but would be irradiated by the neutrons. No activity was detected on this second target. Traces of sodium contamination on the target might also have led to the formation of ^{24}Na, but spectroscopic analysis of the magnesium ribbon used for making the targets failed to show the presence of sodium.

The activity of half-life 2·1 hr. evident in figure 6 was assigned to nucleus ^{18}F, since absorption measurements gave the β^{+} particle energy as 0·92 MeV (Seaborg, 1944, gives 112 min. and 0·7 MeV for ^{18}F). This might have been formed in the reaction ^{17}O (d,n) ^{18}F (Welles 1941) from oxygen contamination on the targets.

Observations on the short-period decay of magnesium targets are shown in figure 7. The observed half-life of 7·0 ± 0·2 sec. was attributed to ^{26}Al formed in accordance with reaction (3). The same period has been found previously in a number of reactions—^{23}Na (α,n) ^{26}Al, ^{25}Mg (p,γ) ^{26}Al, ^{26}Mg (p,n) ^{26}Al, ^{27}Al (γ,n) ^{26}Al —but measurements of the β^{+} energy vary from 1·8 to 4·6 MeV (Frisch 1934; Magnan 1937; White, Delsasso, Fox & Creutz 1939; Curran & Strothers 1939; Huber, Lienhard, Scherrer & Waeffler 1943). The absorption of the β^{+} particles in aluminium was investigated by measuring the short-period β^{+} counting rate after successive bombardments (each lasting 20 sec.) but with different thicknesses of aluminium between the target and the counter; the results (figure 8) give a β^+ energy of 2·8 MeV.

The same decay period and β^+ particle absorption were also obtained by bombarding magnesium with protons (Curran & Strothers 1939), as would be expected from reaction (6). White *et al.* suggest that ^{25}Al may also be produced in these bombardments by the reactions ^{24}Mg$(d,n)^{25}$Al and ^{24}Mg$(p,\gamma)^{25}$Al, and that the ^{25}Al may have a period and β^+ particle energy closely similar to those of ^{26}Al. No complexity is apparent in the curves, however, and the results for the (p,γ) and (d,n) reactions appear identical, though the relative amounts of ^{25}Al and ^{26}Al present might be expected to be different. There is thus no evidence for the formation of ^{25}Al.

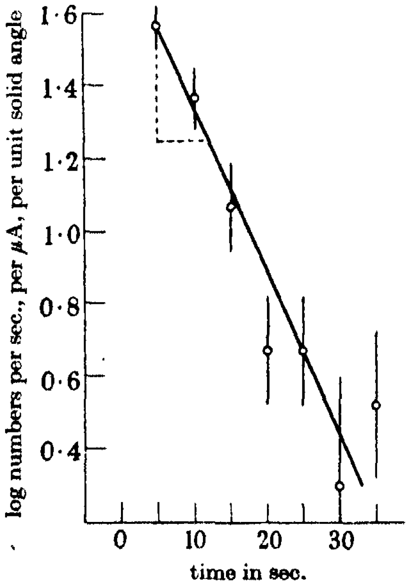


FIGURE 7. Decay curve of ^{26}Al activity.

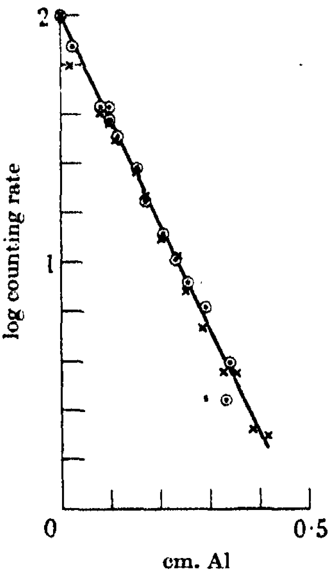


FIGURE 8. Absorption of β^+ particles from ^{26}Al. ⊙, ^{25}Mg (p,γ) ^{26}Al; ×, ^{25}Mg$(d,n)^{26}$Al.

Figure 9 shows the variation of the intensity of the 7 sec. β^+ activity, i.e. of the amount of ^{26}Al formed, on bombarding a magnesium target with deuterons of different energies. The resonance is referred to in § 4 (*b*).

The 20 min. positron activity, found by Ridenour, Delsasso, White & Sherr (1938) after the bombardment of magnesium by 3·8 MeV deuterons was not observed, being masked perhaps by the strong 10 min. positron activity from ^{13}N.

(*b*) *Aluminium*

The 2·4 min. ^{28}Al formed according to reaction (4) was detected in aluminium targets after deuteron bombardment. It was necessary as with magnesium to use a magnet to reduce the background due to positrons from ^{13}N. The rapid decay of the activity while the background was still intense prevented accurate measurements being taken.

The decay in activity of an aluminium target after it had been bombarded for several hours showed the presence of ^{18}F (2·1 hr.) due to oxygen contamination, and also a small amount of ^{24}Na (14·8 hr.) probably due to the (d,p) reaction in sodium contamination. The ^{24}Na activity might have come from the reaction $^{27}\mathrm{Al}\,(n,\alpha)\,^{24}\mathrm{Na}$, the neutrons being produced at the target by a mechanism already discussed for magnesium, but an aluminium plate placed behind the target during the bombardment showed no subsequent activity, and this explanation is therefore unlikely.

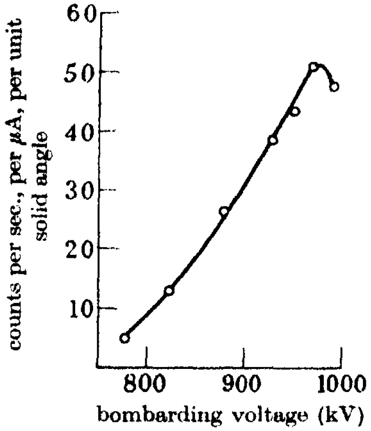


FIGURE 9. Yield of 7 sec. activity.

(*c*) *Silicon*

The silicon target after deuteron bombardment showed the presence of an activity decaying with a half-life of 175 min. as expected from reaction (5). Allowing for decay and geometrical factors the yield of β-particles at the time of ending the bombardment was found to be 260 per sec. This number should also be the number of protons per sec. from the reaction $^{30}\mathrm{Si}\,(d,p)\,^{31}\mathrm{Si}$; but the actual counting rate for the proton group assigned to the reaction is only 20 per sec.

4. DISCUSSION

(*a*) *Assignment of proton groups and mass values*

Table 1, column (3), gives the Q values associated with the proton groups observed in the bombardment of magnesium, aluminium and silicon by deuterons. The isotopic constitution of magnesium and silicon is:

^{24}Mg	77·4 %	^{25}Mg	11·5 %	^{26}Mg	11·1 %	(Mattauch)
^{28}Si	92·28 %	^{29}Si	4·67 %	^{30}Si	3·05 %	(Inghram)

and since each of the proton groups observed in the bombardment may arise from any of the isotopes of that element, the assignment of each to its proper reaction is not easy. In the first announcement of the results of the present work (Allan & Clavier 1946) the Q values to be expected from the possible (d,p) reactions in the magnesium isotopes were calculated from the masses published by Mattauch (1946), and the proton groups were assigned to give the best possible fit with these calculated

Q values. Recent work (Elder, Motz & Davison 1947; Bleuler & Zuenti 1947) has shown that the masses published by Mattauch need revision; also the method becomes ambiguous since the probable errors in the calculated Q values may be of the same order as the differences between them (0·5 MeV). For these reasons the attempt to obtain detailed agreement between observed and calculated Q values has been abandoned; and different assignments of the proton groups have been made, based on the assumption that in a sequence of nuclei of the same isotopic number but with mass numbers differing in multiples of four, the Q values for similar reactions lie on a smooth curve. (See, for example, Barkas 1947.) This method becomes unreliable where there is a discontinuity in the variation of binding energy with atomic number, i.e. where a new nuclear shell is formed. There is, however, no sure evidence for this before mass number 40 in the range considered.

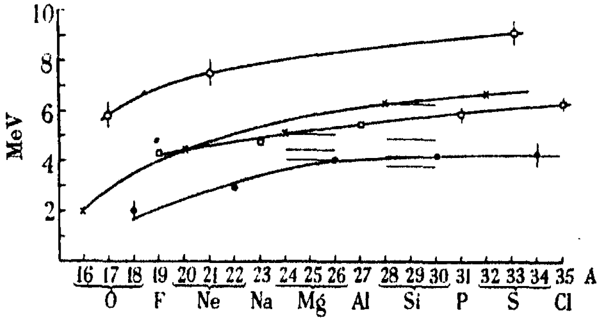


FIGURE 10. Q values for analogous reactions.

×, $4n$; ○, $4n+1$; ●, $4n+2$; □, $4n+3$.

Figure 10 shows the Q values for (d,p) reactions in nuclei of mass numbers $4n$, $4n+1$, $4n+2$ and $4n+3$; except for the values obtained in the present work, and those obtained by Elder, Motz & Davison for neon, all the data are taken from Mattauch's tables. A consideration of the general trend of energies in analogous reactions makes the following assignments appear reasonable:

$^{24}\mathrm{Mg}\,(d,p)\,^{25}\mathrm{Mg}$ $Q = 5\cdot03\,\mathrm{MeV}$ $^{26}\mathrm{Mg}\,(d,p)\,^{27}\mathrm{Mg}$ $Q = 4\cdot05\,\mathrm{MeV}$

$^{28}\mathrm{Si}\,(d,p)\,^{29}\mathrm{Si}$ $Q = 6\cdot16\,\mathrm{MeV}$ $^{30}\mathrm{Si}\,(d,p)\,^{31}\mathrm{Si}$ $Q = 4\cdot16\,\mathrm{MeV}$

None of the observed Q values falls anywhere near the predicted value for the reactions $^{25}\mathrm{Mg}\,(d,p)\,^{26}\mathrm{Mg}$ or $^{29}\mathrm{Si}\,(d,p)\,^{30}\mathrm{Si}$. It may be that the predicted long-range proton groups exist (range about 100 cm.) but have not been detected because they are of low intensity at 930 kV bombarding voltage, or have a non-isotropic angular distribution with a minimum at the 90° angle of observation used in these experiments.†

The Q value of 4·45 MeV found with magnesium could be assigned to the reaction $^{24}\mathrm{Mg}\,(d,p)\,^{25}\mathrm{Mg}$ and if so, would indicate an excited state in the $^{25}\mathrm{Mg}$ nucleus at

---

† [*Added in proof.*] Measurements made recently at 30° and 120° to the deuteron beam have also failed to show the predicted long-range proton group from magnesium targets.

0·58 MeV. The excitation function of the group corresponding to this Q value (figure 5) is, however, different from that for the group corresponding to the transition to the ground state of ^{25}Mg, and is similar to that for the reaction ^{25}Mg (d, n) ^{26}Al. Thus it is perhaps more likely that the 4·45 MeV group arises in the reaction ^{25}Mg (d, p) ^{26}Mg* (see § 4 (b)). If this is so, the difference between the observed and the calculated energy release would indicate that the ^{26}Mg nucleus is left in an excited state at about 4 MeV. Other excited states, at 1·85 and 3·00 MeV were reported by Pollard & Humphreys (1941).

The interpretation of the results for aluminium, which has a single stable isotope, ^{27}Al, presents no difficulty. The maximum energy release of 5·46 MeV fits well with the series of Q values in analogous reactions with ^{19}F (4·3), ^{23}Na (4·76), ^{31}P (5·9), and ^{35}Cl (6·31) and the smaller Q values indicate excited states in the ^{28}Al nucleus at 1·00, 1·48 and 2·15 MeV above the ground state. McMillan & Lawrence (1935) have reported excited states at 0·68, 2·69, 3·67 and 5·15 MeV.

The Q values 3·75 and 4·87 MeV have been assigned to ^{28}Si, thus indicating excited states in the ^{29}Si nucleus at 1·3 and 2·4 MeV, only because ^{28}Si is the most abundant isotope (92·3 %).

On the basis of these assignments, and taking into account the results of other workers, a new set of mass values has been drawn up in table 2. The masses given are derived from two standards, the mass spectrograph value for ^{27}Al, 26·98980 ± 0·00008 (Fluegge & Mattauch 1943), and the mass of ^{24}Mg, 23·9927 ± 0·0004, which is obtained from the mass of ^{23}Na (Elder, Motz & Davison 1947) and two well-known Q values (Murrell & Smith 1939; Siegbahn 1946).

TABLE 2

nucleus	atomic mass	reaction used and Q value	reference
^{25}Mg	24·9939 ± 0·0004	^{24}Mg (d, p) ^{25}Mg 5·03 ± 0·05 MeV	present work
^{27}Mg	26·9926 ± 0·0003	$^{27}\mathrm{Mg} \xrightarrow{\beta+\gamma} {}^{27}\mathrm{Al}$ 2·65 ± 0·25 MeV	Bleuler & Zuenti; Bothe (mean value)
^{26}Mg	25·9904 ± 0·0003	^{26}Mg (d, p) ^{27}Mg 4·05 ± 0·05 MeV	present work
^{28}Al	27·9905 ± 0·0001	^{27}Al (d, p) ^{28}Al 5·46 ± 0·06 MeV	present work
^{28}Si	27·9855 ± 0·0003	$^{28}\mathrm{Al} \xrightarrow{\beta+\gamma} {}^{28}\mathrm{Si}$ 4·6 ± 0·25 MeV	Bleuler & Zuenti; Bothe
^{29}Si	28·9855 ± 0·0003	^{28}Si (d, p) ^{29}Si 6·16 ± 0·06 MeV	present work
^{30}Si	29·9831 ± 0·0001	^{27}Al (α, p) ^{30}Si 2·26 MeV	Livingston & Bethe; Merhaut
^{31}Si	30·9852 ± 0·0001	^{30}Si (d, p) ^{31}Si 4·16 ± 0·06 MeV	present work

From the masses of table 2, the mass of ^{23}Na, and Fluegge & Mattauch's masses for ^{31}P and ^{32}S, the Q values for a number of reactions have been calculated and compared with those obtained by experiments (table 3). The experimental values are all taken from Mattauch's tables (1946) except that for the ^{32}S (n, α) ^{29}Si reaction, which comes from Stebler & Hüber (1948). The agreement between the observed and calculated values is not very satisfactory.†

TABLE 3

reaction	Q value (MeV) calc.	obs.
^{24}Mg (α, p) ^{27}Al	$-1{\cdot}3 \pm 0{\cdot}4$	$-1{\cdot}82$
^{27}Al (d, α) ^{25}Mg	$6{\cdot}3 \pm 0{\cdot}4$	$6{\cdot}46 \pm 0{\cdot}14$
^{23}Na (α, p) ^{26}Mg	$2{\cdot}3 \pm 0{\cdot}4$	$1{\cdot}91$
^{25}Mg (α, p) ^{28}Al	$-0{\cdot}9 \pm 0{\cdot}4$	$-1{\cdot}05$
^{28}Si (α, p) ^{31}P	$-1{\cdot}2 \pm 0{\cdot}4$	$-2{\cdot}23$
^{32}S (n, α) ^{29}Si	$0{\cdot}4 \pm 0{\cdot}4$	$1{\cdot}16 \pm 0{\cdot}15$
$^{31}\mathrm{Si} \xrightarrow{\beta} {}^{31}\mathrm{P}$	$2{\cdot}5 \pm 0{\cdot}3$	$1{\cdot}8$

(b) *Excitation functions and yields*

The thin target excitation functions for the proton groups from magnesium and aluminium (figure 5) all show a rapid increase in yield between 700 and 1000 kV. This is to be expected owing to the increasing probability of penetration of the Coulomb barrier by the incident deuteron; in all cases the protons observed are emitted with an energy greater than the barrier height and the yield of the reaction is therefore governed by the energy of the incident particle.

The ratio of the probability of the (d, α) and (d, p) reactions in ^{26}Mg was estimated from the intensity of the ^{24}Na and ^{27}Mg activities to be 7/1000 at 930 keV deuteron energy. This is consistent with the observations of Henderson (1935) and Bothe (1944) at higher energies; the intensity difference is to be ascribed principally to the different penetrabilities of the potential barrier of the ^{27}Mg and ^{24}Na nuclei for the emitted proton and α-particle. A rough estimate of the ratio of these penetrabilities was made using Bethe's formulae (1937) with the values for the nuclear radii given by Amaldi *et al.* (1947). The result, 1/350, is of the same order as the ratio of the activities.

The excitation function for the group tentatively attributed to the reaction ^{25}Mg (d, p) ^{26}Mg* shows a broad resonance at 955 ± 15 kV. A resonance at 975 ± 15 kV was found in the yield of radioactive ^{26}Al formed in the reaction ^{25}Mg (d, n) ^{26}Al. Both resonances may plausibly be identified with an excited state of the same compound nucleus, in agreement with the above assignment. The compound nucleus ^{27}Al has an excitation energy of at least 17 MeV, and, according to current theory, a continuum of overlapping levels is expected at this excitation. Broad resonances have, however, been observed before in (d, p) reactions (Bennett, Bonner, Hudspeth, Richards & Watt 1941; Martin 1947), and may indicate the operation of selection rules which reduce very considerably the number of levels effective in the particular disintegration.

---

† [*Added in proof.*] A recent determination of the Q value for the reaction ^{27}Al (p, α) ^{24}Mg, $1{\cdot}32 \pm 0{\cdot}06$ MeV (Freeman 1948), is in excellent agreement with the adopted mass values, and gives much smaller probable errors.

This work was carried out in the Cavendish High-Tension Laboratory, Cambridge. The authors wish to acknowledge with sincere thanks the help and guidance given by Dr W. E. Burcham during the course of this work; also the valuable technical assistance of Mr W. Birtwhistle. They wish to express their indebtedness to Trinity College, Cambridge and the Department of Scientific and Industrial Research (H.R.A.), and to the British Council (C.A.W.) for awards made in the period of this research.

REFERENCES

Allan, H. R. & Clavier, C. A. 1946 *Nature*, **158**, 832.

Amaldi, E., Bocciarelli, D., Cacciapuoti, B. N. & Trabacchi, G. C. 1947 *Physical Society: Report of the Conference on Fundamental Particles*, p. 97.

Barkas, W. H. 1947 *Phys. Rev.* **72**, 346.

Bennett, W. E., Bonner, T. W., Hudspeth, E., Richards, H. T. & Watt, B. E. 1941 *Phys. Rev.* **59**, 781.

Bethe, H. A. 1937 *Rev. Mod. Phys.* **9**, 69.

Bleuler, E. & Zuenti, W. 1947 *Helv. Phys. Acta*, **20**, 195.

Bothe, W. 1944 *Z. Phys.* **123**, 1.

Bothe, W. 1946 S.I.G.E.S.O. Supplement no. 3 (Ministry of Supply) 47.

Burcham, W. E. & Smith, C. L. 1938 *Proc. Roy. Soc.* A, **168**, 176.

Cockcroft, J. D. & Walton, E. T. 1934 *Proc. Roy. Soc.* A, **144**, 704.

Curran, S. C. & Strothers, J. E. 1939 *Proc. Roy. Soc.* A, **172**, 72.

Elder, F. K., Motz, H. T. & Davison, P. W. 1947 *Phys. Rev.* **71**, 917.

Fluegge, S. & Mattauch, J. 1943 *Phys. Z.* **44**, 181.

Freeman, J. 1948 (Unpublished.)

Frisch, O. R. 1934 *Nature*, **133**, 721.

Henderson, M. C. 1935 *Phys. Rev.* **48**, 855.

Huber, O., Lienhard, O., Scherrer, P. & Waeffler, H. 1943 *Helv. Phys. Acta*, **16**, 33.

Inghram, M. G. 1946 *Phys. Rev.* **70**, 653.

Livingston, M. S. & Bethe, H. A. 1937 *Rev. Mod. Phys.* **9**, 245.

McMillan, E. & Lawrence, E. O. 1935 *Phys. Rev.* **47**, 343.

Magnan, C. 1937 *C.R. Acad. Sci., Paris*, **205**, 1147.

Martin, A. B. 1947 *Phys. Rev.* **72**, 378.

Mattauch, J. 1946 *Nuclear Physics Tables*. New York: Interscience Publishers Inc.

Merhaut, O. 1940 *Phys. Z.* **41**, 528.

Murrell, E. B. M. & Smith, C. L. 1939 *Proc. Roy. Soc.* A, **173**, 410.

Oliphant, M. L. E., Kempton, A. E. & Lord Rutherford 1935 *Proc. Roy. Soc.* A, **150**, 241.

Pollard, E. 1940 *Phys. Rev.* **57**, 1186.

Pollard, E. & Humphreys, R. F. 1941 *Phys. Rev.* **59**, 466.

Ridenour, L. N., Delsasso, L. A., White, M. G. & Sherr, R. 1938 *Phys. Rev.* **53**, 770.

Schultz, H. L., Davidson, W. L., Jr, & Ott, L. H. 1940 *Phys. Rev.* **58**, 1043.

Seaborg, G. T. 1944 *Rev. Mod. Phys.* **16**, 1.

Siegbahn, K. 1946 *Phys. Rev.* **70**, 127.

Stebler, A. & Huber, P. 1948 *Helv. Phys. Acta*, **21**, 59.

Welles, S. B. 1941 *Phys. Rev.* **59**, 679.

White, M. G., Delsasso, L. A., Fox, J. G. & Creutz, E. C. 1939 *Phys. Rev.* **56**, 512.

Heat transfer by radiation to surfaces at low temperatures

By M. Blackman, Ph.D., Sir Alfred Egerton, Sec.R.S.
and E. V. Truter, Ph.D.

(*Received* 5 *November* 1947)

A study of the transfer of heat between the walls of vacuum vessels is made. It is shown that the heat transferred from the outer wall at ordinary temperature to the inner vessel at 90° K is greater than would be expected from the reflectivity of the inner wall, as estimated from its electrical conductivity. The apparent emissivities at 90° K of copper, silver, gold, tin, brass, aluminium, steel and graphite were determined. It was found that a freshly reduced smooth surface of copper was more efficient in diminishing the heat transfer than a highly polished surface.

1. Introduction

The object of the work here described was to provide information for the design of large vacuum-jacketed vessels for the conveyance of liquefied gases. Precise information was needed about the transfer of heat as it depends on the metal used for the construction of the vessel and on the condition of the surfaces.

Although there have been a considerable number of researches on vacuum-jacketed vessels since the classical work of Dewar, the subject does not seem to have been studied so completely as its practical importance and theoretical interest would seem to justify. Indeed, the conclusions reached as a result of the work here described lead to further experimental work being desirable, for there is still a discrepancy between the theoretical and experimental results.

A brief review will suffice to indicate what had been published on the subject. Although Dewar used a vacuum-jacketed calorimeter in 1873, no account of his researches on vacuum vessels was published until 1893. Dewar (1894) stated that a 30-fold improvement was obtained by evacuating the air from within the jacket and silvering the walls. In 1898 he published a study of the efficiencies of glass vacuum flasks with various coatings on the surfaces and with various substances in the vacuum space. The results obtained were only approximate. Although Dewar & Tait had used charcoal to produce a vacuum in 1874, it was not until 1904 that Dewar drew attention to the use of charcoal cooled in liquid air to give high vacua. This enabled him to develop the metal 'vacuum vessel', but again it was not for many years that he published an account (1919) of his experience in the construction of such vessels.

Banneitz, Rhein & Kurze (1920) studied vessels constructed of glass, plastics and metal; they devised a method of measuring approximately the gain of heat via the neck independently of the other gains, and by using very high vacua, having eliminated conduction through the gas space, they could determine the transfer of heat by radiation across the jacket.

Briggs (1920–1) confirmed that the neck was not responsible for much of the heat reaching the interior of the vessel, because a large part of the heat conducted down the neck is used in warming the cold gas rather than in evaporating the liquid.

Linlithgow Library.

The state of knowledge of the whole subject up to 1923 was summarized in the report of The Oxygen Research Committee of the Department of Scientific and Industrial Research (London), in which the values then accepted for the emissivities of various metals which might be employed in the construction of vessels were included.

Lambert & Hartley (1926) noticed that two similar vessels, one coated with copper and the other with silver, were not equally efficient when used for storing liquid air, although they had been so when storing hot water, and they endeavoured to find the cause of this discrepancy. Four glass vessels were attached to the same pumping line, metallic films having been deposited on to the glass by chemical and electrolytic means. Good agreement was not obtained between the various series of experiments and no explanation of this was given. It was suggested that the emissivity of copper was normal at 373° K, but abnormally high at 90° K.

Meissner & Voigt (1930) examined the efficiencies of vacuum-jacketed vessels made of german silver for the storage of liquid air and liquid hydrogen. These authors made the most complete attempt to deal with the theory of the vacuum vessel.

Roebuck (1943) described the construction of a triple-walled vessel and stated that the evaporation loss was in accordance with theory.

There has been little attempt to study the effect of change in the nature of the metallic surfaces on the radiant heat transfer, and the present work was undertaken to determine emissivities under conditions such as exist in vacuum vessels. It has usually been taken for granted, following Dewar, that the vacuum adjacent surfaces in metal vessels should be as highly polished as possible, but this is not necessarily so, because a highly polished surface is generally covered with a thin oxide film and also retains some of the polishing material.

2. EXPERIMENTAL METHOD

In order to compare the emissivities of various surfaces at low temperatures under conditions similar to those existing in vacuum vessels, it was necessary to devise a simple apparatus in which substitution of the various surfaces could be easily achieved, yet providing as closely comparable results as possible. Figure 1 indicates the type and dimensions of the apparatus which was used.

Two identical cylindrical vessels, constructed of pyrex glass, fitted with large external ground joints, are connected to the same vacuum system. The caps of the joints carry the inner cylindrical vessels. The long narrow neck of the inner vessel is surrounded by a concentric jacket of the same diameter as the inner vessel. Both the inner vessel and its neck jacket can be filled separately with liquid oxygen. The rate at which heat is received by the inner vessel can be measured by the rate of evaporation of the liquid oxygen within it. The liquid oxygen in the neck jackets absorbs any heat which would otherwise enter by the neck. The area receiving the radiant heat conveyed to the inner vessel is also rendered constant by the existence of the neck jacket. The surfaces of the inner or outer vessels can be chemically coated with metals, or metal foils can be placed around the inner vessel. The pressure of residual gas in the vacuum space is the same for both vessels: a McLeod and a Phillip's gauge were included in the vacuum system. The vessels were immersed in a large thermostat of water maintained at 20 ± 0·1° C. Two pairs of such vessels were often in use at the same time. The rate of evaporation of the liquid oxygen in the inner vessel was determined by a capillary flowmeter. Flow was made even by including between the vessels and the flowmeter a 10 l. vessel containing a little water, which served to saturate the oxygen.

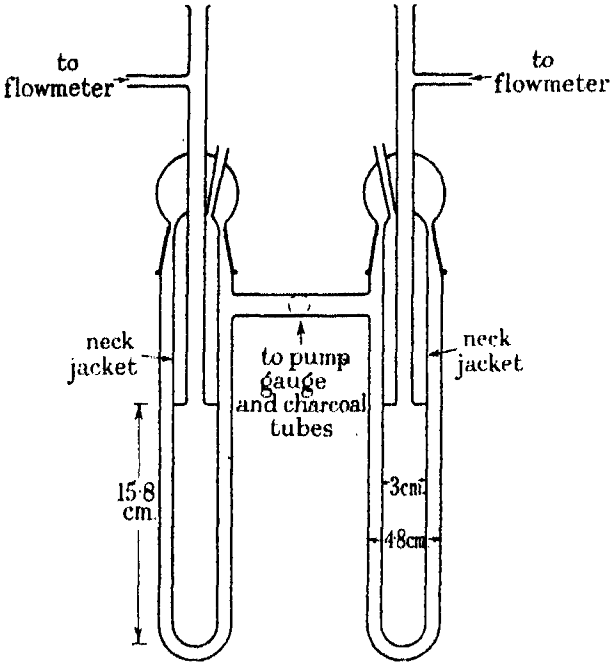


FIGURE 1. Vacuum vessels.

Having described the main features of the apparatus, some of the experimental details are now given under the headings to which they refer; discussion of the experimental errors will be found after the theoretical section (§ 4). Many preliminary experiments were needed to establish a satisfactory procedure.

(*a*) *The metallic surfaces*

Silver was deposited by chemical reduction on either the inner surface of the outer vessel, or the outer surface of the inner vessel, or both. Brashear's silvering process was used; a uniform deposit tenacious enough to polish and free from pinholes was obtained. For other metals, and for silver too, foil usually well polished was wrapped round the outer surface of the inner vessel and held in position by binding with fine copper wire. The foils were cut to be the same area (135·5 cm.^{2}) and placed in similar positions on the inner vessel. Silver was chemically deposited on the surface of the inner vessel before affixing the foil. The hemispherical end of the inner vessel was thus left silvered; the heat absorbed by the silver end was allowed for when estimating the heat transfer through the foil. Since no appreciable difference was found between the results with polished and unpolished chemically deposited silver, this underlying deposit was not polished, for fear of any damage to the film; if any small area of the hemispherical end is unsilvered, large errors are caused.

Experiments showed that even without the silvering, the contact of the foils with the outer surface of the inner vessel was sufficiently good to give results equivalent to those which would have been obtained if the metal had been in perfect contact with the glass: whether four or eight bindings were used, the result was the same. Assuming the area of contact is only equivalent to strips round the inner vessel of width equal to the diameter of the binding wire, it can be shown that the conductivity of the metal (steel for instance) and of the glass is sufficient to prevent a temperature difference between the two faces of the glass wall of the inner vessel greater than 3·5° C; this temperature difference would only lead to a difference of 0·4 % in the heat transferred. The following foils were used:

Silver ('fine') Brass (64 % copper) Cadmium Aluminium Austenitic steel (F.S.L. and I.M. 5[*]) Copper (rolled)	These foils varied in thickness from 0·012 to 0·05 cm.
Gold (fine) Tin ('pure')	0·0025 cm. thick.
Chromium	An electrolytic deposit (0·005 cm. thick) on copper foil.

[*] These were of commercial purity supplied by the late Dr Hatfield, F.R.S.

The silver, copper and brass foils were polished using rouge. The finish was a fine, almost optical, polish with a few fine scratches. Cadmium, aluminium, chromium and F.S.L. steel were polished on the buffing wheel without the use of any polishing agents. The finish of the gold and tin foils, as obtained from Messrs Johnson, Matthey, was about equal to that of the buffed surfaces. After polishing or buffing, the surfaces were washed several times with distilled ether, free from any grease.

Certain of the foils referred to in § 3 received special treatment. The 'matt reduced copper' surfaces were prepared by heating the foil in air until it had oxidized and then plunging it into methyl alcohol while still hot. The foil was then transferred to the test vessel as rapidly as possible. The 'polished reduced copper surfaces' were prepared by reducing the surface with hydrogen. After the surface had been polished and degreased, it was placed in an electrically heated evacuated tube. Hydrogen was then admitted into the tube up to slightly less than atmospheric pressure, and the temperature was controlled at 330° C. After 5 min. the contents of the tube were allowed to cool to room temperature in an atmosphere of hydrogen. In the later experiments, reduction was carried out *in situ* at 305° C by boiling benzophenone in the inner vessel by means of an immersion heater, and admitting hydrogen to 500 mm. pressure into the vacuum jacket. After a reduction of about 5 min. at 305° C, evacuation was started and the vessel was baked out at this temperature for some 5 or 6 hr. This procedure ensured that the metal surface was free from grease except in so far as such contamination may have occurred during the determination of the evaporation rate by vapours distilled on to the cold surface, for which there is no evidence.

The graphite was painted on from a water suspension and polished when dry.

(*b*) *The vacuum*

Two pairs of vessels (see figure 2) were connected to the same vacuum line. One of the vessels was kept unchanged during a series of experiments and the results obtained served as a check on the other measurements. Apeizon 'L' grease was used on all ground joints and all the large vacuum stopcocks. A mercury two-stage diffusion pump with liquid air trap backed by a rotary oil pump was used for the preliminary evacuation. One of the larger pyrex tubes containing charcoal was employed as a 'getter' for maintaining a high vacuum. The preliminary evacuation was continued while the inner vessel and charcoal tubes were heated by steam until the vacuum was less than 10^{-5} mm. on the Phillip's gauge. This usually took 48 hr. The pump was then shut off and the charcoal tube cooled in liquid air. There was then, of course, no reading on the gauge and the vacuum was probably of the order of 10^{-6} mm. In the final experiments on copper surfaces, the inner vessel was heated to 305° C, hydrogen being introduced to reduce the copper. The subsequent evacuation was necessarily still more thorough.

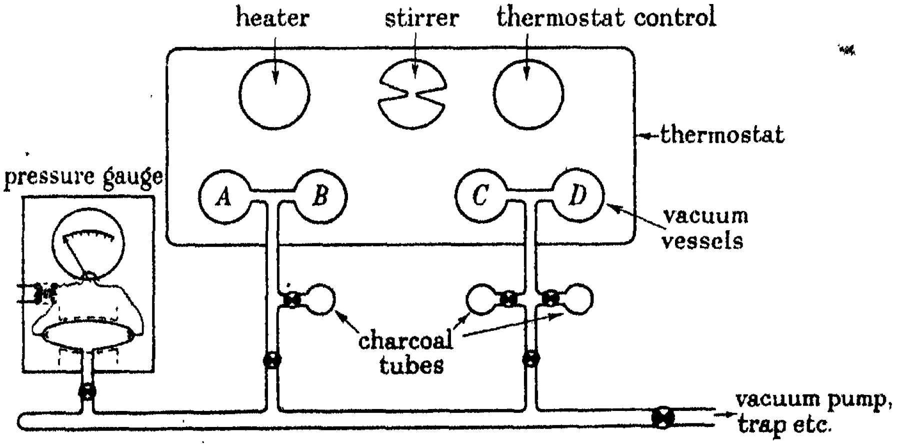


FIGURE 2. Arrangement of apparatus.

The conduction of heat through the residual gas in the vacuum space was the same for all readings in each set of experiments, but the pressure of the gas, though not higher than 10^{-5} mm. was not infinitely low, and therefore experiments were carried out to test the error introduced by conduction through the residual gas. The evaporation rates for various pressures, as measured by a McLeod gauge, are shown in table 1.

A comparison of the evaporation rates at 10^{-5} mm. and zero pressure gave a measure of the heat transferred by the residual gas: the difference was 0·5 ml./min., the corresponding quantity of heat being $0{\cdot}4 \times 10^{-5}$ cal./cm.^{2} sec. (cf. Soddy & Berry 1910, 1911).

TABLE 1

pressure (mm. × 10^{-4})	evaporation rate (ml.(S.T.P.)/min.)
11·5	82·5
6·0	56·0
4·0	43·6
0·9	29·1
0·5	26·2

(c) *Measurement of evaporation rate*

The flowmeter had a range 0 to 120 ml./min.: the calibration was made with oxygen at eleven different flow rates, the probable error of the curve of best fit was 0·4 ml./min. The size of capillary was chosen so that conditions of streamline flow held throughout.

The possibility of the superheating of the liquid oxygen was tested from time to time, but with a thermocouple sensitive to 0·3° C, no such superheating was ever detected.

Preliminary experiments showed that after filling the vessel a period of at least 2 hr. was necessary before a steady state was completely established.

The method of measuring the evaporation rate was as follows: Fifty consecutive 10 sec. readings of the flowmeter were taken from a fixed mark very low down on the neck of the inner vessel 0·5 cm. above the bottom of the neck jacket: the vessel was then refilled. When the level reached the mark again another fifty readings were taken. The error in metering the gas was taken as the root mean square of the deviations of the readings from the average, viz. about 1·5 %. There was a slight decrease in the evaporation with falling level. The lowest rate of evaporation during an experiment was about 12 ml./min., which corresponds to the transfer of only about 1 cal./min. since the latent heat of vaporization of oxygen is 51·0 cal./g., 1 ml. (S.T.P.)/min. $\sim 8{\cdot}15 \times 10^{-6}$ cal./cm.^{2} sec.

3. EXPERIMENTAL RESULTS

The experimental observations are summarized in table 2.

Each value in table 2 is the average of at least two determinations in two different vessels. For the combinations of surfaces shown in table 3 many more values were obtained. The bracketed values were for the same unaltered surface. They provide a guide as to the extent of the error inherent in the technique which is 6 or 7 %.

There were four sets of experiments in which comparison of matt with polished surfaces were made (table 4).

These metals are arranged in order of increasing grain size of the matt surface. The matt silver and rolled copper surfaces were only slightly more dull than the corresponding polished ones; the 'matt and reduced copper' surface (reduced after oxidation) was very coarse and in places it was possible to cause pieces of metal to flake off; the matt steel had a grain size intermediate between these two extremes.

The general conclusion that can be drawn from these experiments is that for a double-walled metallic vessel (in which visible radiation is completely eliminated) a matt surface is as efficient as a polished one, provided that the grain size of the matt surface is not too large and that the polishing process does not change the properties of the surface. It is noteworthy that the polished copper surface is slightly less efficient than the rolled foil surface which has not been polished.

TABLE 2

hot wall (293° K)	cold wall	evaporation rate. Volume of oxygen at S.T.P. (ml./min.)
glass	'reduced copper'	21·5 ± 2·0
	chemical silver	25·5 ± 2·9
	gold	27·6
	'rolled copper'	33·5
	'matt and reduced copper'	34·7
	copper (rolled and polished)	38·9
	silver	39·5 ± 3·0
	tin	41·7
	brass	50·9
	cadmium	59·6
	aluminium	60·5
	chromium	71·6
	steel (F.S.L.)	109·2
	matt steel (F.S.L.)	117
	graphite	546
	glass	941
chemical silver	glass	61·2
	chemical silver	17·1 ± 1·6

The observations show that the surface that absorbs least heat at 90° K is polished copper which is then reduced. The rate of evaporation is nearly one-half that given by a polished copper surface which is not so reduced. This improvement is to be expected from the fact that copper in air is always covered with a layer of cuprous oxide. Preston & Bircumshaw (1935) examined various metallic surfaces by means of electron diffraction and estimated that a freshly reduced copper surface would acquire a layer of cuprous oxide about 250×10^{-8} cm. thick in 1 hr. at 27° C and cuprous oxide is known to show strong absorption in the infra-red.

A suitable surface can be obtained by reducing a rolled surface at 300° C taking care to prevent unnecessary oxidation before the reduction is carried out. The vessel has to be made with welded or hard-soldered joints. The baking out at 300° C also ensures a better final vacuum as compared with the general practice of baking out at 100° C.

The chemically deposited silver was about 50 % more efficient as a reflecting surface than was fine silver foil. Careful examination of the silver foil surface revealed that it had a very faint pink tint as though covered with a very fine layer of rouge dust. This tint could not be removed by vigorous polishing with chamois leather. Silver is known to be covered with a superficial layer of oxide after exposure to air.

TABLE 3

glass/silver foil (ml./min.)		glass/chemical silver (ml./min.)	chemical silver/ chemical silver (ml./min.)
31·1		32·8 *	15·0
32·7		34·0	16·4
39·0	39·1 ± 2·7 (7 %)	23·1	19·2
41·0		26·0	16·5
32·3		24·6	20·0
31·3		22·9	16·4
43·2		24·6	17·6
42·7		25·6	15·6
43·5		24·0	19·1
40·3		26·5	13·7
40·0		20·4	15·3
38·1		20·8	14·0
47·3	42·2 ± 2·6 (6 %)	21·2	21·3
43·0		20·5	19·7
42·1		30·6	
36·8		30·3	
45·4			
39·0			
37·0			
37·7			
42·1			
42·9			
39·5 ± 3·0 (8 %)		25·5 ± 2·9 (11 %)	17·1 ± 1·6 (9 %)

* This higher value was probably due to slight imperfections in the deposit, but has been included in assessing the mean value.

chemical silver/glass (ml./min.)	glass/polished and reduced copper (ml./min.)	glass/gold (ml./min.)
59·0	24·2	27·8
62·2	25·7	27·1
60·5	19·3 *	26·0
65·1	20·2	27·5
64·0	20·5 *	29·0
63·2	19·4	28·0
56·1		
59·4	21·5 ± 2·0 (9 %)	27·6
61·2 ± 2·0		

* Reduced *in situ*.

TABLE 4

metal	$\frac{\text{evaporation from matt}}{\text{evaporation from polished}}$ surface
chemically deposited silver	1
copper (rolled)	0·86
steel (F.S.L.)	1·08
reduced copper	1·61

The chemical silver was deposited by a reducing process when required and should, therefore, have a much thinner layer of oxide on the surface. Moreover, only the foil had been polished with rouge. The polishing process produces an amorphous layer on the foil, whereas the deposition process probably produces a fine crystalline surface. Other experiments have shown that the difference between the value for the foil and for the chemically deposited silver is not due to the difference in contact with the surface of the inner vessel. These results are further discussed in § 6.

4. THEORETICAL

Attempts have been made by several authors to determine the heat flow, by radiation, across the vacuum space of a vacuum vessel from theoretical considerations. As the theory is presented here on the same lines as advanced by Meissner & Voigt (1930), only the main steps will be given. The difference between Meissner's presentation and this, lies in the methods used to obtain the values of the integrated quantities.

Consider a vacuum to be bounded by two infinitely large parallel plates at different temperatures. At equilibrium the energy (W) transferred from the plate at the higher temperature to that at the lower is made up of two parts: (a) That emitted by plate 1 and absorbed by plate 2, less (b) that emitted by plate 2 and absorbed by plate 1. To render the calculation less laborious, each plate is assumed to emit only monochromatic radiation corresponding to that of maximum intensity calculated from the Planck equation. Then

$$W = \sigma T'^4 \int_0^{\frac{1}{2}\pi} \left\{\frac{A_\perp A'_\perp}{1-R_\perp R'_\perp} + \frac{A_\parallel A'_\parallel}{1-R_\parallel R'_\parallel}\right\}_{\lambda'} \cos\theta \sin\theta \, d\theta - \sigma T^4 \times \int_0^{\frac{1}{2}\pi} \left\{\frac{A'_\perp A_\perp}{1-R_\perp R'_\perp} + \frac{A'_\parallel A_\parallel}{1-R_\parallel R'_\parallel}\right\}_{\lambda} \cos\theta \sin\theta \, d\theta, \qquad (1)$$

where θ is the angle to the normal, A and R are the absorptivity and reflectivity respectively and the subscripts $\perp$ and $\parallel$ refer to plane polarized rays perpendicular to or parallel with the plane containing the incident and reflected rays.

In this equation the total energy transfer is made up of two functions which relate to different wave bands. Consequently the values of A and R referring to the wavelength λ are not the same as those referring to λ'. Nevertheless, to simplify the calculations the two emissivity factors are assumed to be equal. This introduces an error of only 1 %. The equation then becomes

$$W = \sigma(T'^4 - T^4) \int_0^{\frac{1}{2}\pi} \left\{\frac{A'_\perp A_\perp}{1-R'_\perp R_\perp} + \frac{A'_\parallel A_\parallel}{1-R'_\parallel R_\parallel}\right\}_{\lambda'} \sin\theta \cos\theta \, d\theta. \qquad (2)$$

Maxwell first deduced the relation between the optical constants of a metal and its reflectivity. The relation between the optical constants and the electrical conductivity was obtained by Drude (1894). By combining these relations the relation between absorptivity and the electrical conductivity κ is found to be

$$A_\parallel = \frac{4\sqrt{(\kappa\tau)}\cos\theta}{[\sqrt{(\kappa\tau)}+\cos\theta]^2+\kappa\tau}, \quad A_\perp = \frac{4\sqrt{(\kappa\tau)}\,(\cos\theta)^{-1}}{(\sqrt{(\kappa\tau)}+(\cos\theta)^{-1})^2+\kappa\tau},$$

where τ is the period of vibration of the radiation.

Values of $A_\perp$ and $A_\parallel$ are calculated for various values of θ from 0 to $\frac{1}{2}\pi$, using the value of τ for the incident radiation (viz. $3{\cdot}32 \times 10^{-14}$ sec.) and of the electrical conductivity κ for the material at the temperature of the wall. Substituting these values of A into equation (2), a graph of the function $dW/d\theta$ is plotted against θ and the area under the curve measured using an Amsler planimeter.

The absorption and reflexion constants for silica at normal incidence for wavelengths between 5 and 20μ were used for calculating the reflecting power of the glass surface at various angles. The reflecting power for each angle was then averaged over the Planck function corresponding to the temperature 293° K. Except near glancing incidence, $80° < \theta < 90°$, the absorptivity of glass is large compared with that of metals consequently the heat absorbed by the *cold* surface of a glass-metal pair is almost independent of the actual value of the absorptivities for glass, and so only a small error is introduced by using the known values for silica instead of those for Pyrex, which are similar but have not been accurately determined.

The specific resistance is sensitive to the presence of traces of impurities, and these impurities cause a greater percentage error at 90 than at 293° K. With brass, steel, aluminium and tin, it was possible to measure the resistance of the actual foil used in the heat transfer experiments; for copper and silver it was preferable to measure the specific resistance of a commercial product of the same quality as the foil, rather than to take figures from the literature, which are for pure materials. However, for cadmium, chromium and gold this had to be done.

A precision Callendar & Griffiths bridge was used for the resistance measurements, which were made at 90, 293 and 393° K, and recorded in table 5.

TABLE 5

	specific resistance ρ in c.g.s. units $\times 10^{-6}$			
metal	373° K	273° K	90° K	293° K
tin	17·25	12·28	4·20	—
alloy steel	131·1	117·3	86·6	—
brass	8·22	7·18	5·17	—
aluminium	3·87	2·743	0·519	—
copper	2·273	1·60	0·323	—
silver	2·15	1·568	0·445	1·684

Other values of $\rho_{90°}$ used were (Gruneisen, *Handb. Exp. Phys.*):

Gold	$0{\cdot}567 \times 10^{-6}$ c.g.s. unit.
Chromium	$1{\cdot}82 \times 10^{-6}$ c.g.s. unit.
Cadmium	$1{\cdot}98 \times 10^{-6}$ c.g.s. unit.

Two typical graphs (figures 3 and 4) are given in illustration of the results; emissivities dE are plotted against θ, obtained from

$$dE = \left\{\frac{A'_\parallel A_\parallel}{A'_\parallel + A_\parallel - A'_\parallel A_\parallel} + \frac{A'_\perp A_\perp}{A'_\perp + A_\perp - A'_\perp A_\perp}\right\} \sin\theta \cos\theta \, d\theta, \tag{3}$$

in which the measured specific resistances have been used.

Figure 4 illustrates the special case silver/silver where one of the silver surfaces is at 293° K; the effect of angle of incidence is evident.

Next, some of the limitations to the application of the theory will be considered.

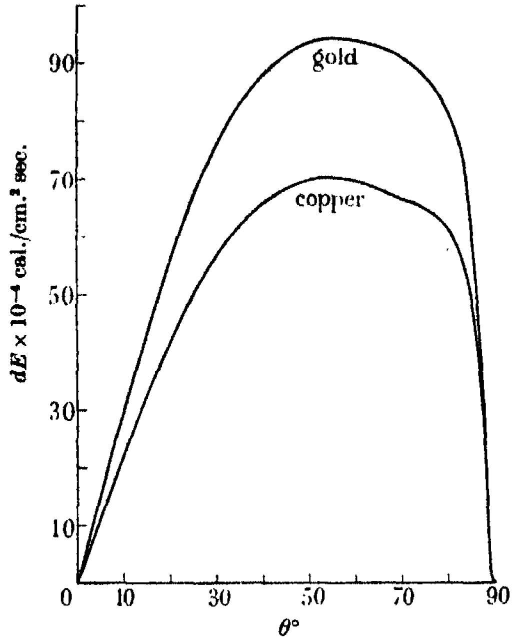


FIGURE 3

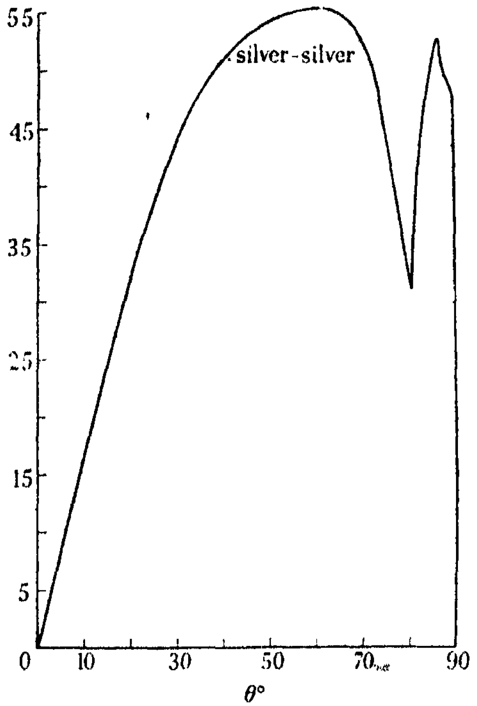


FIGURE 4

(*a*) *Correction for the shape of vessel*

The heat transfer by radiation across the jacket of a vacuum vessel cannot be considered precisely as that between two parallel planes at different temperatures: the vessels are of various shapes and sizes.

For concentric spheres the problem is tractable, as the paths of the incident and reflected rays can be determined. In other cases the geometry can become troublesome and the estimation of 'end-effects' becomes difficult, especially when the inner and outer vessel have appreciably different areas. The following theorem, however, helps to simplify matters; consider two concentric spherical vessels at the same temperature, having a vacuum between them and having different emissivities (which may be a function of the angle of emission), since the temperature of the two vessels remains constant it follows that the heat absorbed by the inner vessel is equal to that emitted by it. Hence calculation of the latter may be used to evaluate the former. Since the only factors involved are the emission as a function of angle and the shapes and relative positions of the vessels, the geometrical part of the calculation will be valid even if the temperatures of the vessels are different. Thus in order to calculate the energy absorbed by the inner vessel, the energy emitted by it can be estimated, inserting for the temperature the value for the outer vessel.

For a cylindrical vessel the fraction of the radiation from the outer wall which does not reach the inner is an important fraction of the total, if the areas of the inner and outer walls are very different. In order to evaluate the heat transfer in the experimental vessels a calculation is first made for concentric infinite cylinders made of silver with radii equal to those of the experimental vessels. The emissivities of the two walls are calculated from their electrical conductivity with respect to the wave-length pertaining to the higher temperature (i.e. $9{\cdot}96\mu$). The rays emitted from the inner vessel are characterized, as usual, by the angles ϕ and ψ, ϕ being the angle from the normal and ψ the azimuthal angle, the zero of which is parallel to the common axes of the cylinders. For a given value of ϕ, the angle of incidence on the outer cylinder θ is calculated for various values of ψ. The successive reflexions of this ray then proceed at the same angles of incidence (ϕ and θ), and the total radiation (for ψ, ϕ) is then evaluated in the same way as for parallel planes. The calculation is then repeated for other values of ϕ and the final result obtained numerically. This calculation leads to the value $5{\cdot}8 \times 10^{-5}$ cal./cm.^{2} sec. for the radiant energy transfer between the two silver cylinders.

For the actual vessel, a modification has to be made in the calculation to allow for the cylinders being finite, it is assumed: (1) that all the radiation from the base is lost, and (2) that half the radiation from the 'edge' is lost; the 'edge' being taken as a distance equal to the difference in radii. With this correction the value of the radiation transfer becomes $6{\cdot}3 \times 10^{-5}$ cal./cm.^{2} sec.

When the outer wall was glass and the inner was metal, the variation of the absorptivity of glass is neglected because the absorption is so large. For the same reason nothing is gained by carrying out a lengthy calculation on the exact contributions by successive reflexions. The calculation is carried out for parallel planes and the result is corrected for the 'end-effect' as described above; thus, because of the geometry, no more than 2 % need be added to the heat transfer calculated for the glass to silver surfaces; the same conclusions hold when the inner wall is glass and the outer is metal.

(b) *Correction for emissivity change with temperature and contribution of the cold wall*

Only a small error in the heat transfer is introduced by neglecting the contribution of the cold wall to the radiation from the hot wall. The value of $\lambda_{\text{max.}}$ corresponding to the hot wall is less than that corresponding to the cold wall and the absorptivity of metals will decrease as the wave-length increases. Therefore, if the heat transfer is calculated by neglecting the energy radiated by the cold wall, and again by assuming that the two emissivity factors are equal, then the true heat transfer will be somewhere between these two limits. These limits differ by 1 % and therefore the error introduced by using either approximation will not be serious.

(c) *Correction for assumption of monochromatic radiation*

It has been assumed that each surface has emitted only monochromatic radiation, corresponding to that for which the emissive power of a black body at the same temperature is a maximum. The basis of the calculation was that

$$W = \sigma(T_1^4 - T_2^4)\,F, \tag{4}$$

where from equation (2)

$$F = \int_0^{\frac{1}{2}\pi} \left\{ \frac{A_\parallel A'_\parallel}{A_\parallel + A'_\parallel - A_\parallel A'_\parallel} + \frac{A_\perp A'_\perp}{A_\perp + A'_\perp - A_\perp A'_\perp} \right\}_\tau \sin\theta \cos\theta \, d\theta.$$

Here A and A' refer to the two surfaces neither of which is black, and τ has the value corresponding to $\lambda_{\max}$. But F is really a function of λ and therefore this procedure is not strictly justified. The energy emitted, expressed more accurately, is

$$W = \int_0^\infty K_\lambda F_\lambda \, d\lambda, \tag{5}$$

where F_λ is the value of F corresponding to the wave-length λ and K_λ denotes the emissive power for λ.

A calculation of the values of W (equation (5)) over the range $\lambda = 3$ to 60μ was made using Planck's equation and integrating graphically and the result was compared with the value of W obtained from equation (4).

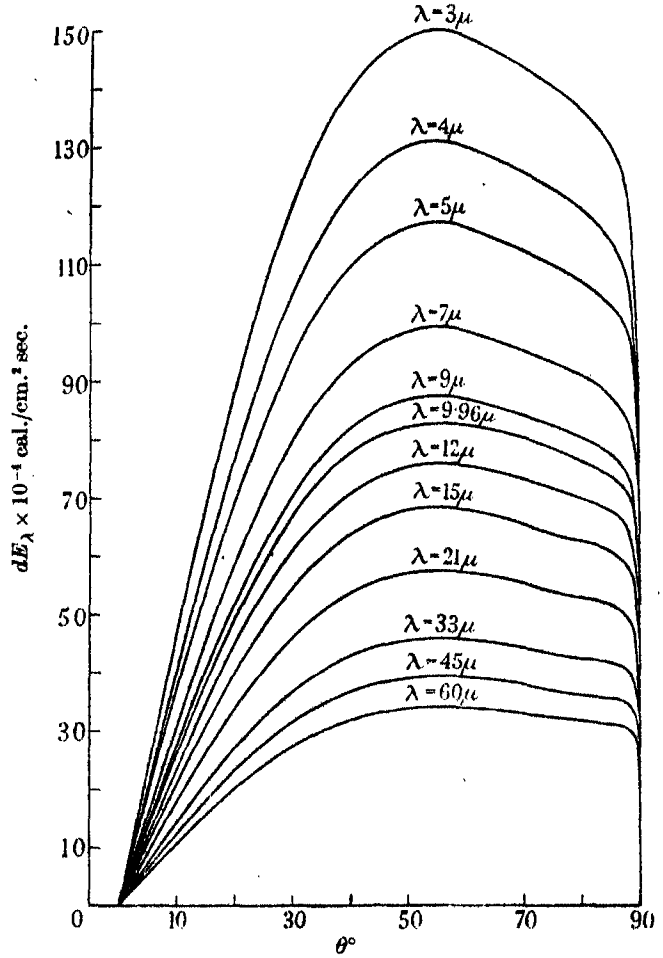


FIGURE 5. $dE_\lambda = \left\{\left(\frac{4\sqrt{(\kappa\tau)}\cos\theta}{[\sqrt{(\kappa\tau)}+\cos\theta]^2+\kappa\tau}\right)+\left(\frac{4\sqrt{(\kappa\tau)}(\cos\theta)^{-1}}{(\sqrt{(\kappa\tau)}+(\cos\theta)^{-1})^2+\kappa\tau}\right)\right\}_\lambda \sin\theta\cos\theta\,d\theta.$

The graphs (figures 5 and 6) show calculations for the vessel with black surface (293° K)/silver (90° K). The values obtained for the two methods of calculation were:

(1) Using Planck integral for values from 3 to 60μ, $8\cdot5 \times 10^{-5}$ cal./cm.^{2} sec.

(2) Approximate calculation using $\lambda_{\max.} = 9\cdot96 \times 10^{-4}$ cm. and multiplying by σT^4, $10\cdot1 \times 10^{-5}$ cal./cm.^{2} sec.

Therefore the deviation of the more rigorous from the approximate estimate is 16 %.

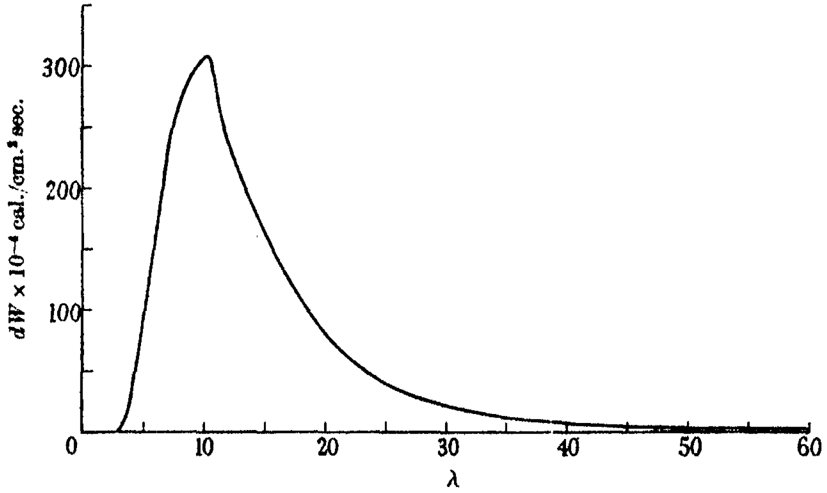


FIGURE 6. $dW = \frac{2\pi hc^2}{\lambda^5} \cdot \frac{1}{e^{ch/\lambda kT} - 1} E_\lambda d\lambda$.

Without these three corrections (*a*), (*b*) and (*c*), the theoretical values are probably estimated about 15 to 20 % too high. It is well to have established, in view of what follows later in the discussion, that the calculated values are the maximum rather than the most probable.

5. ASSESSMENT OF EXPERIMENTAL ERROR

The experimental vessel is not perfect in that heat can enter other than by radiation from the outer jacket to the inner vessel, and it is necessary to assess the quantities of heat which can otherwise enter.

(*a*) The conduction of heat through the residual gas has already been considered (see table 1) and does not exceed about 0·4 ml./min.

(*b*) The heat entering down the narrow neck by radiation from the stopper and reaching the liquid oxygen in the inner vessel can be readily estimated.

Assuming that the stopper and the liquid oxygen are black bodies, the radiant heat transfer (cf. equation (9)) will be given by

$$\sigma(T_1^4 - T_2^4)\,\pi r\{(z^2 + r^2)^{\frac{1}{2}} - z\}, \tag{6}$$

where r is the radius of the neck (0·55 cm.), z is the difference in level between the oxygen in the neck jacket and in the neck (about 8 cm.). From this expression it follows that the evaporation loss from this cause should only be 0·3 ml./min.

(c) The fact that the experimental vessel consisted of two finite concentric cylinders necessitated a correction for the entry of radiant heat into the ends. The cylinders terminated in a pair of hemispheres; and no correction is necessary at that end. At the other end, the cylinders terminated at a level where the neck jacket began. When the outer wall was glass about as much energy would be radiated into the vessel by the glass wall opposite the neck jacket as would be radiated out, and so no correction is then necessary. When, however, the outer wall was covered with a deposit of silver up to a certain height, then by virtue of its much smaller emissivity more energy would be radiated into the vessel from the glass above it than would be radiated out by the silvered wall of the vessel, and the net result would be an increase in the evaporation rate. A rather elaborate investigation was needed to determine the magnitude of this 'end-effect'.

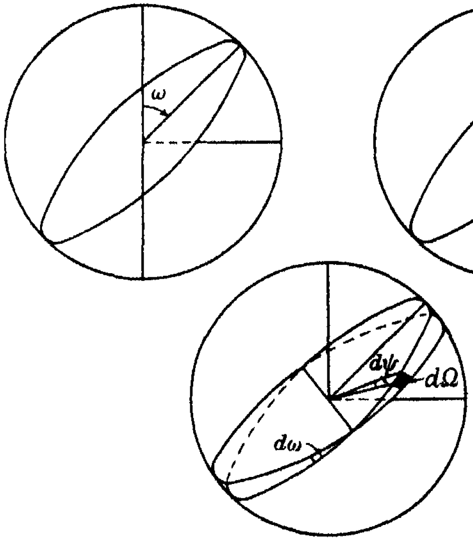


FIGURE 7

The calculation of this magnitude under various circumstances necessitated using co-ordinates (see figure 7) defined as follows:

$$\cos\theta = \cos\psi\cos\omega \quad \text{and} \quad d\Omega = \cos\psi\, d\psi\, d\omega.$$

These co-ordinates are applied to the experimental vessel, as shown in figure 8.

The energy radiated from the area $dS = 2\pi r_0 dz$ into the region below $z = 0$ is

$$E\sigma T^4 \frac{1}{\pi}\iint \cos\theta\, d\Omega\, ds \tag{7}$$

$$= E\sigma T^4 \frac{1}{\pi}\int_{\psi=-\frac{1}{2}\pi}^{\psi=\frac{1}{2}\pi}\int_{\omega=\omega_1}^{\omega=\frac{1}{2}\pi} \cos^2\psi\, d\psi\, \cos\omega\, d\omega\, 2\pi r_0\, dz$$

$$= E\sigma T^4\, 2r_0\, dz\{1-\sin\omega_1\}\, \tfrac{1}{2}\pi, \tag{8}$$

but $\sin\omega_1 = z/(z^2+a^2)^{\frac{1}{2}}$, a being the width of the annulus, therefore the total radiation R into the region below $z = 0$ from the region above $z = z_1$ is

$$E\sigma T^4 \pi r_0 \int_{z=z_1}^{z=\infty} \left\{1 - \frac{z}{(z^2+a^2)^{\frac{1}{2}}}\right\} dz,$$

therefore

$$R = E\sigma T^4 \pi r_0 \{(z^2+a^2)^{\frac{1}{2}} - z\}. \tag{9}$$

This expression (9) represents the energy radiated into the vessel from the remainder of the wall of the outer vessel above a distance z cm. higher than the level of a horizontal plane passing through the position where the neck jacket begins (see figure 8). Inserting the values

$$r = 2\cdot15\,\text{cm.},\quad a = 0\cdot65\,\text{cm.},\quad z = 0\,\text{cm.},\quad E = 0\cdot86,$$

the total energy radiated into the upper or 'open' end of the cylinder by the outer glass wall becomes

$R = 3\cdot77 \times 10^{-2}$ cal./sec. (corresponding to evaporation = 31·0 ml./min.).

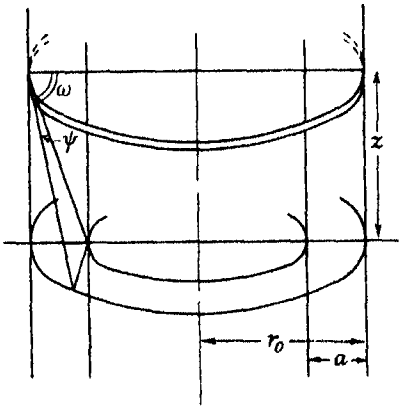


FIGURE 8

Consider the two combinations of surface

(i) Silver (293° K), Glass (90° K).
(ii) Silver (293° K), Silver (90° K).

(i) *Silver/glass.* After the first experiment with this type of vessel, care was taken to ensure that there was always about 1·5 cm. of silvering above the level $z = 0$, i.e. $z_1 > 1\cdot5$ cm. Glass has a high emissivity so that very little of the radiation incident on the inner vessel will be reflected and about 98 % will be returned from the silver surface, most of which will be absorbed on its second incidence. The inner vessel can therefore be considered to have an emissivity of unity.

If the energy radiated into the vessel, measured in terms of oxygen evaporated is 31·0 ml./min. for $z = 0$ as above, because $z \sim 1\cdot5$ (figure 9), only 20 % will be absorbed by the inner surface, i.e. 6·2 ml./min. This constitutes about 10 % of the total evaporation observed (61 ml./min.) and may therefore be detectable experimentally.

The evaporation rate for a certain vessel was measured and the particular 'end-effect' calculated 5·8 ml./min. Then using the same surfaces a copper cone was placed around the neck jacket and adjusted so that the bottom of the cone was level with the top of the inner vessel, and the diameter of the base of the cone was about 2 mm. less than that of the outer vessel. This arrangement was designed to eliminate the 'end-effect' almost completely:

Evaporation without cone	63·3 ml./min.
End correction	−5·8
	57·5 ml./min.
Measured evaporation with cone	57·8 ml./min.

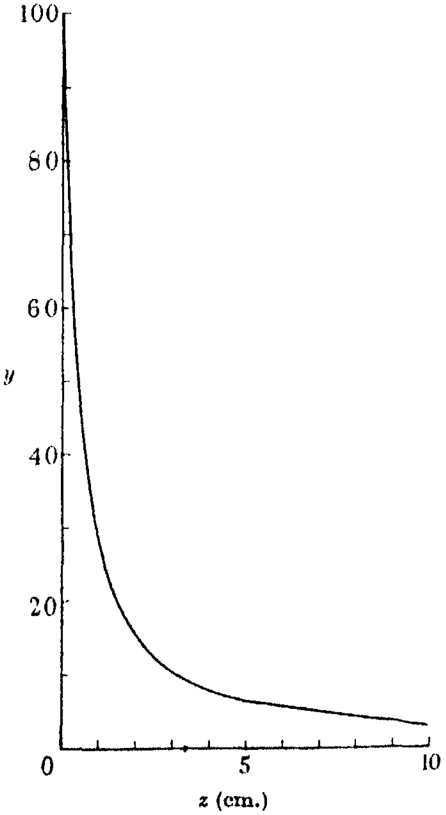


FIGURE 9. $y = 100a/\{(z^2+a^2)^{\frac{1}{2}}-z\}$.

This close agreement may be fortuitous because the experimental error is of the same order as the difference observed. Although not accurately measured in every experiment the probable value of z was not so different from that assumed, as to give rise to an error greater than 3 ml./min.

(ii) *Silver/silver.* The two surfaces were normally coated by the silvering process to the same height; in some instances, however, the upper part of the deposit on the neck jacket was removed so that z had positive values. The evaporation from both types of vessels was such that no difference could be detected experimentally between them, which indicates that this 'end-effect' is less than the experimental error.

A rather elaborate analysis was made to estimate the energy absorbed by the inner vessel due to radiation entering from the end, the value 2×10^{-5} cal./cm.^{2} sec. being obtained which is of the same order as the experimental errors and cannot therefore be detected.

When one of the experimental vessels was immersed in liquid oxygen, being inserted to about the height of the top of the inner vessel, heat can presumably only enter by what has been referred to above as the 'end-effect' and by radiation down the neck itself, which together would account for an evaporation 0·6 ml./min., leaving an unaccounted evaporation of 1·2 ml./min. No allowance has been made for this small possible influx of heat in the final assessment of the errors.

(*d*) There is another source of error to be considered; the liquid oxygen in the neck jacket boiled freely at atmospheric pressure, while the liquid in the inner vessel was boiling under a slightly increased pressure due to the flowmeter and the saturator vessel. For an evaporation of 25 ml./min., the difference of boiling-point was estimated to amount to about 0·12° C and this would cause a slight difference in the transfer of heat via the neck according to the height of the liquid oxygen in the neck. Experiments were made to test the extent of the effect, but the results showed that it was at the limit of the observable and it was only when the pressure was purposely and considerably increased that an effect was measurable.

(*e*) *Summary of experimental errors.* The various sources of error which have been discussed may now be summarized (see table 6).

TABLE 6

cause	error (ml./min.)
flowmeter: (*a*) observational	± 0·4
(*b*) instrumental	± 0·4
conduction by residual gas*	± 0·4
radiation down neck*	± 0·3
total	± 1·5

* Maximum error rather than the most probable.

Correction amounting to not more than + 10 % has also to be made for the effect of the difference in boiling-point between the oxygen in the neck jacket and the inner vessel. There is also the possible unaccounted error of about 1·2 ml./min. For metal/glass and metal/metal surfaces there are 'end-corrections' estimated to be about − 3 and − 2·2 ml./min. respectively. The errors estimated above are of the same order as the experimental error in table 3, which is about 10 %. From this examination, the total absolute error should not exceed 20 % of the measured heat transfer under any circumstances.

6. DISCUSSION OF RESULTS

In table 7 are set out in column 3 the values for the heat transfer from the hot wall to the cold inner vessel as found experimentally. The values are corrected in accordance with the discussion in the previous section. The theoretical values estimated according to the methods in § 4 are given in column 4. Column 5 gives the ratio of the values in column 3 to those in column 4, and column 6 gives values for the emissivities as determined from the experiments.

TABLE 7

1 hot wall (293° K ± 0·1)	2 cold wall (90° ± 0·7)	3 experimental values (corrected)	4 theoretical values	5 ratio col. 3/col. 4	6 emissivity
		$\times 10^{-5}$ cal./cm.^{2} sec.			
glass	polished copper (reduced)	20·3	8·1	2·5	0·019
	chemical silver	23·6	9·6	2·5	0·023
	gold	25·4	10·9	2·3	0·026
	rolled copper	30·6	8·1	3·8	—
	matt copper (reduced)	31·7	8·1	3·9	0·032
	polished copper	35·2	8·1	4·4	0·035
	silver	35·8	9·6	3·8	0·036
	tin	37·8	27·9	1·4	0·038
	brass	45·7	30·7	1·5	0·046
	cadmium	53·5	19·5	2·7	0·054
	aluminium	54·3	10·4	5·2	0·055
	chromium	64·0	19·1	3·4	0·065
	steel (F.S.L.)	95·2	75·6	1·3	0·096
	steel matt	101·9	…	—	—
	graphite	444	—	—	0·48
	glass	765	—	—	0·87
chemical silver	glass	52·5	17·3	3·0	—
	chemical silver	14·9	6·3	2·4	—
			(6·5 for parallel planes)	mean 2·9	

It is clear from this table that the theoretically derived values of heat transfer are much smaller than the measured ones; except in three cases (tin, brass and steel) there can be no doubt of this significant deviation. The greatest deviation found was for aluminium.

As mentioned in the introduction all other experimental work carried out in this field confirms this conclusion that the efficiency of a metal as a reflecting surface in a vacuum vessel cannot be estimated precisely from its electrical conductivity. For instance, the results of Banneitz *et al.* (1920) make the radiation loss for a silvered vessel of 2 l. capacity $10{\cdot}1 \times 10^{-5}$ cal./cm.^{2} sec., whereas theoretically it should have been $6{\cdot}5 \times 10^{-5}$ cal./cm.^{2} sec. The following values are got from a study of Lambert & Hartley's results having regard to the dimensions of their apparatus:

TABLE 8

	experimental ($\times 10^{-5}$ cal./cm.^{2}/sec.)	theory	$\frac{\text{experimental}}{\text{theory}}$
glass/Ag	29·2	9·6	3·0
Au	30·7	10·9	2·8
Pt	45·2	—	—
Cu	66·8	8·1	8·2
glass/Ag	22·7	9·6	2·4
Ag/glass	35·2	17·3	2·0
Ag/Ag	19·0	6·5	2·9

Meissner & Voigt's results are summarized in table 9:

TABLE 9

	theory			experiment (averaged values)
vessel	neck	radiation	total	(cal./cm.^{2} sec.)
iron/german silver (80° K)	$9{\cdot}1 \times 10^{-5}$	$36{\cdot}6 \times 10^{-5}$	$45{\cdot}7 \times 10^{-5}$	$33{\cdot}1 \times 10^{-5}$
silver/silver (80° K)	$9{\cdot}1 \times 10^{-5}$	$6{\cdot}5 \times 10^{-5}$	$15{\cdot}6 \times 10^{-5}$	$23{\cdot}6 \times 10^{-5}$
silver/silver (21° K)	$11{\cdot}8 \times 10^{-5}$	$1{\cdot}9 \times 10^{-5}$	$13{\cdot}7 \times 10^{-5}$	$21{\cdot}9 \times 10^{-5}$

The divergence between the values for iron/german silver (in which the experimental value is less than the theoretical) was attributed to the fact that the theoretical calculation was approximate. In the other cases the silver surface was supposed to be contaminated. In the two experiments using liquid air, the neck loss must have been too high because no allowance for the heat absorbed by the issuing gas appears to have been made. Even so, when this theoretical neck loss is subtracted from the experimental loss then the difference which is a measure of the radiation loss is still not in agreement with the theoretical figure.

TABLE 10

	theoretical radiation	experimental radiation	ratio
Fe/german silver	36·6	24·0	0·66
Ag/Ag	6·5	14·5	2·23

Since the emissivities of the metal surfaces in vacuum vessels do not appear to be calculable from the electrical conductivities, values of the emissivities, as estimated in the following way from the heat transfer experiments, are given in column 6 of table 7.

The thermodynamic relation for the energy exchange between two black bodies is

$$W = \sigma(T_1^4 - T_2^4).$$

If the bodies are not black, but emit only a fraction E of that energy, then by summing the energy exchange the relation

$$W = \sigma(T_1^4 - T_2^4)\frac{E_1 E_2}{E_1 + E_2 - E_1 E_2}$$

can be used to estimate the emissivities.

Since the emissivity of non-metals does not change appreciably with temperature, the value for glass can be obtained from

$$W = \sigma(T_1^4 - T_2^4) \times \frac{E}{2-E} = 765 \times 10^{-5}\,\text{cal./cm.}^2\,\text{sec.}$$

Therefore $$E = 0{\cdot}867.$$

Using this value for glass, the values for E for the metals, in glass/metal pairs were calculated.

As a check on the self-consistency of the results, the two values of the emissivity of silver (at 90 and 293° K) obtained from observations from glass/silver pairs can be combined giving $16{\cdot}7 \times 10^{-5}$ in satisfactory agreement with the experimental value $14{\cdot}9 \times 10^{-5}$ cal./cm.^{2} sec. for a silver/silver pair.

The relation between electrical conductivity and emissivity was established experimentally by the work of Hagen & Rubens (1900, 1902, 1903); this work was not accurate enough to establish the relation with certainty for the highly reflecting metals. More accurate measurements were published later (Hagen & Rubens 1909, 1910). In this later work the energy emitted from a number of metals and alloys was compared directly with that from a black body, at temperatures from 100 to 500° C, and for various wave-lengths 6 to 8, 8·85 and 26μ. The emissivity measured at normal emission agreed moderately well with the calculated value, although there were systematic deviations of about 5 % at 26μ, and more than 20 % at 8·65 and $6{\cdot}65\mu$. These deviations were all in the sense that the emissivity measured was less than that calculated, and in general the deviation was greater when the absolute value of the emissivity was small (e.g. silver (200 to 500° C) $E_{\text{calc.}}/E_{\text{obs.}} = 1{\cdot}68$ but for constantan $E_{\text{calc.}}/E_{\text{obs.}} = 1{\cdot}09$).

Hurst (1933) has carried out similar measurements in the short-wave infra-red region (1 to 5μ), and he also found that the observed emissivity is about 20 % lower than that calculated from the Hagen-Rubens relation, for $\lambda = 5\mu$. It was found impossible to polish copper with rouge without introducing impurities into the surface.

The emissivity measured in all the above cases is for normal emission. In vacuum vessels, however, the emission at all angles plays an important part, and little work has been done to check the theoretical reflexion formulae in the infra-red. Czerny (1924) has, however, measured the variation of emissivity for platinum (at 650° C) with angle of emission for the two types of polarized radiation ($\lambda = 6{\cdot}75\mu$). The measurements were in agreement with the values of the optical constants deduced from the Maxwell theory. Thus there is in all this experimental work a deviation in the opposite direction to that of the heat transfer measurements described in this paper.

Hagen & Rubens used temperatures above 100° C in order that there might be enough energy emitted to be easily measured. At these elevated temperatures films of grease, etc., would be absent and the relative effect of thin films of material, the properties of which differ from those of the bulk of the metal, would be less than at a lower temperature, because of the higher resistivity of the metal itself. Moreover the absorbing effect of thin layers increases with the obliquity of the incident radiation.

The resistivity of a metal is essentially a property of the bulk of the material whereas the absorptivity is a surface property; the properties of these two regions of the metal are not identical. The difference may be partly caused by disturbances in the surface layers of the metal due to polishing leading to an oxide layer or to an amorphous layer, or to both.

(*a*) *Oxide layer.* The effects of an oxide layer can to some extent be examined by comparing the results obtained before and after reducing the surface (see table 7):

	polished copper	reduced polished copper
$\dfrac{\text{experimental heat transfer}}{\text{theoretical heat transfer}}$	4·4	2·5

Here reduction of the surface makes a great difference but does not bring about complete agreement indicating that the oxide layer can account for only part of the discrepancy. The measurements on metals that do not oxidize so readily, e.g. gold (ratio = 2·3), indicate this also.

(*b*) *Amorphous layer*. The thickness of an amorphous layer increases with time of polishing. A rolled copper foil was found in the present work to be not less reflective for heat than a polished one. It is generally recognized that it is almost impossible to reproduce a polished metal surface giving the same optical constants even by the most careful repetition of conditions of polishing (cf. Wheeler 1913). A polished metal surface possesses chemical and mechanical properties which differ from those of the crystalline unstrained metal, and any kind of cold working tends to increase its electrical resistance. Polishing may therefore produce a surface that has properties that are not truly characteristic of the metal itself, quite apart from contamination of the surface by the abrasive or by possible oxidation or corrosion (cf. Lowery *et al.* 1932, 1935, 1936).

There seems little doubt from the present investigation that the surface films, particularly oxide films, influence the emissive powers of the metallic surfaces and account for part of the discrepancy between theory and experiment. It is also possible that the electrical conductivity of the surface of a pure metal crystal is not the same as that of the bulk of the metal, and that this difference shows up at low temperatures and possibly also at low angles of incidence. Another possible explanation of part of the observed discrepancies may be that when radiant energy is received by a cold surface, the gas still retained in the surface layers of the metal influences the heat transfer without affecting the pressure. Extension of this investigation to other ranges of temperature will help to settle these questions.

The authors are indebted to the Ministry of Fuel and Power for a grant for investigations on the use of liquid methane as a fuel, which enabled this work to be brought to the present stage.

REFERENCES

Banneitz, F., Rhein, G. & Kurze, B. 1920 *Ann. Phys., Lpz.*, **61**, 113.
Briggs, H. 1920–1 *Proc. Roy. Soc. Edinb.* **41**, 97.
Czerny, M. 1924 *Z. Phys.* **26**, 182.
Dewar, J. 1893 *Not. Proc. Roy. Instn*, **14**, 1.
Dewar, J. 1894 *Proc. Roy. Soc.* **14**, 393.
Dewar, J. 1898 *Not. Proc. Roy. Instn*, **15**, 815.
Dewar, J. 1904 *Proc. Roy. Soc.* **74**, 127.
Dewar, J. 1919 *Not. Proc. Roy. Instn*, **21**, 591.
Dewar, J. & Tait, W. 1874 *Proc. Roy. Soc. Edinb.* **8**, 348.
Drude, P. 1894 *Physik des Aethers*, 1st ed., p. 574. Stuttgart: Enke.
Grüneisen, E. 1928 *Handb. Exp. Phys.* B, **18**, 164 (ed. Wien, W. & Harms, F.).
Hagen, E. & Rubens, H. 1900 *Ann. Phys., Lpz.*, **1**, 352.
Hagen, E. & Rubens, H. 1902 *Ann. Phys., Lpz.*, **8**, 1.
Hagen, E. & Rubens, H. 1903 *Ann. Phys., Lpz.*, **11**, 873.
Hagen, E. & Rubens, H. 1909 *S.B. preuss. Akad. Wiss.* **1**, 478.
Hagen, E. & Rubens, H. 1910 *S.B. preuss. Akad. Wiss.* **1**, 467.
Hurst, C. 1933 *Proc. Roy. Soc.* A, **142**, 466.

Lambert, B. & Hartley, K. 1926 *Proc. Roy. Soc.* A, **112**, 136.
Lowery, H. & Moore, R. 1932 *Phil. Mag.* **13**, 938.
Lowery, H., Bor, J. & Wilkinson, H. 1935 *Phil. Mag.* **20**, 390.
Lowery, H., Wilkinson, H. & Smare, D. 1936 *Phil. Mag.* **22**, 768.
Meissner, W. & Voigt, B. 1930 *Z. InstrumKde*, **50**, 121.
Preston, G. & Bircumshaw, L. 1935 *Phil. Mag.* **20**, 706.
Roebuck, J. 1943 *Rev. Sci. Instrum.* **14**, 90.
Soddy, F. & Berry, A. J. 1910 *Proc. Roy. Soc.* A, **83**, 254.
Soddy, F. & Berry, A. J. 1911 *Proc. Roy. Soc.* A, **84**, 576.
Wheeler, L. P. 1913 *Amer. J. Sci.* **35**, 493.
The Report of The Oxygen Research Committee 1923 H.M.S.O.

Spectroscopic studies of low-pressure flames; temperature measurements in acetylene flames

BY A. G. GAYDON AND H. G. WOLFHARD, *Imperial College, London*

(*Communicated by Sir Alfred Egerton, F.R.S.—Received* 9 *January* 1948)

Flames at very low pressure have a relatively thick reaction zone (or flame front) and are especially suitable for detailed study of the combustion processes and of the distribution of energy during the reaction. Temperature measurements have been made, by various spectroscopic methods, on flames of acetylene with air, oxygen and nitrous oxide, in some cases down to a pressure of 1·5 mm. Hg. The excitation temperature has been measured through the reaction zone by the spectrum-line reversal method using Fe, introduced as the carbonyl; the characteristics of the flame containing Fe(CO)$_5$ are described. The rotational temperature of the excited OH radicals has been determined from the emission spectrum; at pressures above 10 mm. it is fairly constant at around 5700° K, but at lower pressure rises to a higher value of nearly 9000° K. The results are explained in terms of the collision and radiative deactivation of the electronically excited OH radicals. These radicals are believed to be formed, as the result of chemical reaction, in the excited $^2\Sigma$ state and with a rotational energy equivalent to above 9000° K. Deactivation by collision appears to occur on the average after about forty collisions, if a normal collision diameter is assumed. Removal of electronic excitation occurs mainly by collisions with O$_2$ molecules, but CO$_2$ or CO molecules are most effective in removing rotational energy. The variation of concentration of OH through the reaction zone has been determined by its absorption spectrum; it is abnormally high just below the visible reaction zone. Calculations of flame temperature and composition are given. The lack of equipartition of energy is discussed.

INTRODUCTION

The flat, relatively thick, flames of gases burning at low pressure are particularly suitable for detailed examination of the chemical and physical processes occurring in the reaction zone. Preliminary accounts have already been given of the nature of the phenomena which have been observed in these low-pressure flames (Wolfhard 1943; Gaydon 1947; Gaydon & Wolfhard 1948). The pre-mixed gases are burnt at the top of a quartz or Pyrex tube in a large vessel which is continuously evacuated. It is found that the lowest pressure at which a stable flame can be maintained depends on the diameter of the burner tube, being roughly inversely proportional to this diameter. Thus to obtain flames at very low pressure it is necessary to use tubes of large diameters, and consequently a high mass-flow of gas and high pumping speed.

The acetylene flame is one of the most suitable for investigation, as it is relatively bright and can be run at lower pressure than other gases for similar burner size; with a burner of 35 mm. diameter a flame can be maintained at a pressure slightly below 2 mm. Hg. A previous paper (Klaukens & Wolfhard 1948) has reported some thermocouple measurements on the temperature distribution through this flame, and in this paper temperature measurements by spectroscopic methods are described. It should, of course, be realized that neither thermal nor chemical equilibrium is established in the reaction zone of the flame, and it is not therefore possible strictly to define the temperature. It is, however, possible in some cases to determine an effective temperature for the energy in some particular degree of freedom, and this will enable us to obtain a fuller understanding of the distribution of energy among its various possible forms in the flame gases, and to follow the change in this distribution as we pass through the flame front.

EXCITATION TEMPERATURES BY THE SPECTRUM-LINE REVERSAL METHOD

Flame temperatures are frequently measured by the sodium-line reversal method (for a description and criticism see Gaydon 1948). At atmospheric pressure it is probably a satisfactory method for the main body of the flame, although in the inner cone the emission of the sodium lines is often abnormally strong because the atoms are excited by collisions with active species formed in the combustion processes so that the effective temperature of the sodium atoms exceeds that for thermal equilibrium. We thus arrive at the concept of an 'excitation temperature' in the reaction zone of a flame. This excitation temperature may be defined as equal to the brightness temperature of a background source which, when viewed through the flame, gives the spectrum line under examination neither as a bright line nor reversed by absorption. The value of the excitation temperature may vary with the nature of the element being excited, and even from line to line in the spectrum of the same element.

The usual substance for measuring temperatures by the spectrum-line reversal method is sodium, but other alkali-metals or alkaline earth metals have frequently been used. In low-pressure flames it is difficult to colour the flame with any of these elements because ordinary atomizers will not work at such low pressure. Some preliminary observations were made with the mercury line 2537 A, but absorption in the gases surrounding the flame and in the air of the laboratory caused trouble. One of the simplest metals to introduce into a flame is iron, which can be added as the volatile carbonyl. This substance is an anti-knock, and might influence the flame speed and combustion processes if added in appreciable amount, but in the small concentrations used in these experiments it did not appear to influence the flame speed appreciably.

The flame containing iron carbonyl

The low-pressure flame containing $Fe(CO)_5$ has a rather unusual appearance, and it is of interest briefly to describe the principal features. A diagram of the flame is shown in figure 1. Directly below the main visible flame front is a yellow layer from which yellow streaks spread upwards. There is also a faint yellow cone, slowly changing to green, above the bright blue-green reaction zone of the flame. This yellow region shows the banded spectrum of FeO. Below the reaction zone, and separated from it by a distance about equal to the thickness of the zone, is a pale luminous region, the spectrum of which appears to consist of a continuum from the yellow to blue-green. This 'fore-zone' is distinctly separated from the main reaction zone, and is best seen when the main flame is weakened by adding excess air or O_2. It is presumably connected with the initial decomposition of the carbonyl.

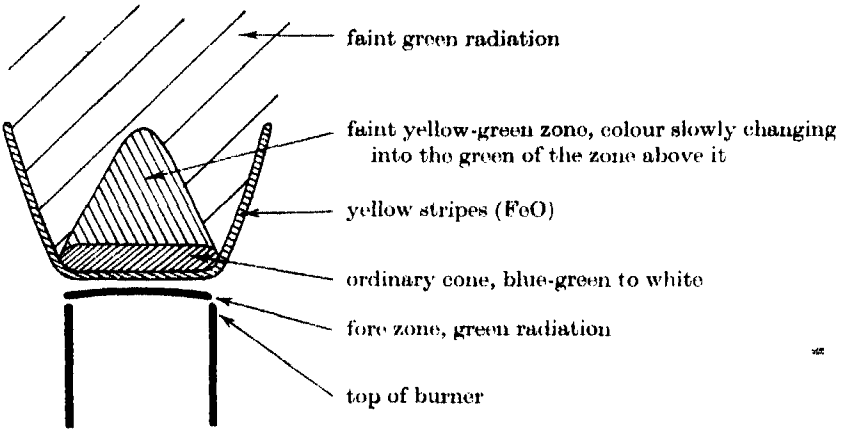

faint green radiation
faint yellow-green zone, colour slowly changing into the green of the zone above it
yellow stripes (FeO)
ordinary cone, blue-green to white
fore zone, green radiation
top of burner

FIGURE 1. Flame coloured with $Fe(CO)_5$.

The experimental arrangement

The excitation temperature of the iron lines was found to be so high that it was found difficult to obtain a black body at sufficiently high temperature to give reversal of the iron lines. With the oxy-acetylene flame, indeed, reversal could not be obtained, and hence the method was only applied to the acetylene-air flame. The hottest available background source was the carbon arc (~ 3800° K), but this did not prove suitable because of some banded structure in its spectrum around 2500 A (CO Fourth Positive and Mulliken's C_2 bands), and because of intense scattered visible light. A hydrogen discharge tube was therefore used as source; in the region of 2500 A this had a brightness temperature of about 3500° K. The hydrogen lamp was of the low-tension hot-cathode type, this having a larger luminous area and giving more uniform illumination than some high-tension types. An image of the hydrogen lamp was formed in the reaction zone of the flame, and a second image of this and the flame together was then formed with a second lens on the slit of the spectrograph. A medium-size quartz spectrograph (Zeiss Q 24) was used. A rotating sector enabled the effective brightness temperature of the hydrogen lamp to be varied at will. The value of this temperature was determined photographically by comparison with a carbon arc, which was assumed to have a brightness temperature of 3855° K in this region. The flame was held as steady as possible throughout the exposure time, and a hair line across the slit and the image of the top of the burner tube enabled the point at which the excitation temperature was being measured to be located with respect to the top of the burner and the emission from the flame. As reported in previous publications, the OH and C_2 bands occur rather lower in the flame than the bands of CH, and it is of interest to compare the excitation temperature of Fe in the flame with the regions of excitation of C_2, etc., as indicated in figure 2. The measurements were made with a burner of 21 mm. diameter, the acetylene-air flame having a pressure of 57 mm. Hg.

Results

While the absolute value of the excitation temperature of the Fe atoms may not be determined with high accuracy, the results are sufficient to show that the excitation temperature greatly exceeds the theoretical flame temperature (see later) and to give the trend of the excitation temperature as we pass through the flame front. The results are shown diagrammatically in figure 2. The temperature obviously must rise from below, and the fall after passing through the centre of the reaction zone is to be attributed to the closer approach to thermal equilibrium as higher points in the flame are reached. All the iron lines do not reverse at exactly the same temperature, and on some plates it is possible to see the Fe line at 2483·28 Å reversed while several neighbouring lines are visible in emission. Our results, showing an abnormally high excitation temperature in these low-pressure flames, differ from the observations by Lewis & von Elbe (1943) using the sodium-line method on flames at atmospheric pressure. The thickness of the reaction zone in the flames at atmospheric pressure is so small that it is doubtful whether their observations were sufficiently sensitive to detect the sharp rise of temperature in this zone.

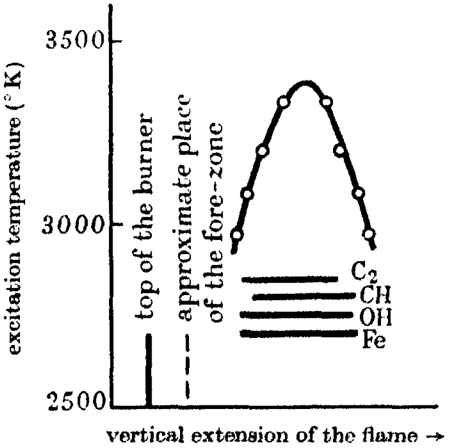


FIGURE 2. Excitation temperature in a C_2H_2-air flame coloured with $Fe(CO)_5$. The positions in the flame at which the various bands and the Fe lines are emitted are indicated. Gas pressure 57 mm. Hg.

ROTATIONAL TEMPERATURE OF THE OH BAND IN EMISSION

The average energy of rotation of the excited OH radicals can be determined from the rotational intensity distribution within the (0, 0) band at 3064 A. Assuming that there is approximately a Maxwell-Boltzmann distribution of rotational energy among the radicals, the result may be expressed as an effective rotational temperature of these radicals. Measurements of this type have been made by a number of investigators (e.g. Oldenberg 1934), and Wolfhard (1939) made some measurements on the inner cones of flames at atmospheric pressure. An ordinary Bunsen flame gave a temperature of 5700° K, while an acetylene-air flame gave about 5400° K. The oxy-acetylene flame gave about 4000° K, the lower value being attributed to mixing of the radiation with that from the hot gases surrounding the inner cone, the concentration of OH being much higher in this flame. The accuracy of these measurements at atmospheric pressure was not very high because of the limited resolving power of the spectrograph employed, but was quite adequate to indicate the order of temperature and to show the lack of equipartition of energy in the flame front. Kondratjew & Ziskin (1937) have made measurements of the rotational temperature of OH in absorption.

Method

Most of the measurements were made using a large Littrow-type quartz spectrograph designed by Dr H. J. Hübner. This has a focal length of about 6 ft. and employs one 60° and one 30° prism. It is especially suitable for intensity work as it was designed to minimize scattered light; the real image of the slit formed by reflexion in the lens is eliminated in the usual way with a stop, while the virtual image is removed by rotating the lens slightly. This instrument has a resolving power of about 50,000, and the Fe lines at 3099·90 and 3099·97 A are clearly resolvable. A few measurements, especially of the weaker flames with air, were made with a medium quartz spectrograph, and the results with the two instruments were consistent to within the accuracy of the measurements, about 300° C. With the smaller instrument the lines Q_1 $6\frac{1}{2}$, $9\frac{1}{2}$, $14\frac{1}{2}$, $15\frac{1}{2}$, Q_2 $6\frac{1}{2}$, $7\frac{1}{2}$, $11\frac{1}{2}$, $12\frac{1}{2}$, $16\frac{1}{2}$, $17\frac{1}{2}$, P_1 $13\frac{1}{2}$, $14\frac{1}{2}$, $15\frac{1}{2}$, $21\frac{1}{2}$, $22\frac{1}{2}$, $23\frac{1}{2}$, $24\frac{1}{2}$, $26\frac{1}{2}$, P_2 $13\frac{1}{2}$, $21\frac{1}{2}$, $23\frac{1}{2}$ and $25\frac{1}{2}$ were found to be free from overlapping and therefore most suitable. With the larger instrument it was possible to use a greater number of lines, including R_2 $16\frac{1}{2}$, $19\frac{1}{2}$ and $23\frac{1}{2}$.

The spectrum of the flame was first photographed using a sharply focused image on the slit; a quartz fluorite achromatic lens was used. A hair across the slit served to give a reference mark along the spectrum to ensure that measurements of all lines were made for the same part of the flame. The plate was then racked down and intensity marks were put on using the usual sector arrangement; for this a lens immediately over the slit was used to focus an image of the source (either an Fe arc or the flame itself) on to the lens of the spectrograph to ensure uniform illumination of the slit. The relative intensities of lines in different parts of the spectrum which have the same initial levels were compared theoretically and experimentally; this enabled errors due to change of plate sensitivity with wave-length or to change of aperture of the spectrograph with wave-length to be checked. Slight errors of this type were found, but correction of these did not significantly change the slope of the line from which the temperature was determined. The intensity measurements were made with the usual type of non-recording microphotometer, using a photocell and galvanometer.

The intensity of any line of the OH band depends on the product of the population of the initial energy level and the transition probability.* Formulae for the transition probabilities for the lines of the various branches of a $^2\Sigma \rightarrow {}^2\Pi$ transition have been given by Earls (1935), and numerical values for the required lines have been evaluated by us using a value of -7 for the constant $\lambda = A/B$. Indicating the transition probability for any line as P, the expression for the intensity I takes the form

$$I = CP\nu^4 e^{-E_r/kT} \approx CP\nu^4 e^{-B'K'(K'+1)/kT},$$

where C is an arbitrary constant, ν is the frequency of the line, T is the absolute temperature and E_r is the rotational energy, which is nearly equal to $B'K'(K'+1)$, where B' is the rotational energy constant and K' is the rotational quantum number for the excited $^2\Sigma$ state. B' has the value 16·95 cm.$^{-1}$. Thus by plotting the log of the observed intensity minus $\log(P \times \nu^4)$ against the rotational energy term, or with sufficient accuracy against $K'(K'+1)$, it is possible to calculate the temperature from the slope of the line. This line will only be straight if the molecules have a Maxwell-Boltzmann distribution of rotational energy. Any failure in this respect, due, for instance, to chemiluminescent excitation to selected levels, will show as a deviation of the curve from a straight line.

Results

The variation of rotational temperature through the reaction zone. For flames of acetylene with oxygen the curve comes fairly straight, indicating a fairly good approach to a Maxwell-Boltzmann distribution of rotational energy. The temperature has been determined at the base, centre and top of the reaction zone, using a burner of 21 mm. diameter and a gas pressure of 6 mm. The results are

(1)	Green zone at base of flame	6200° K.
(2)	The middle of the visible flame	6000° K.
(3)	The blue-violet corona at the top	5300° K.

This indicates that the temperature is highest in the very front of the flame, and falls as higher levels in the flame are reached. The relative total strength of the OH band in these three regions is about 5:2:1.

Variation of temperature with pressure. Using approximately stoichiometric mixtures of oxygen and acetylene, the rotational temperatures have been determined for the centre of the visible flame using burners of different diameter, and therefore different gas pressure.

burner diam. (mm.)	gas pressure (mm. Hg)	rotational temp. (° K)
13	13	5700
21	5·5	6200
34	2·5	7000
52	1·5	8750

---

* The term 'transition probability' as used here includes the effect of differing statistical weights of the levels.

It may be noted that the temperature of 5700° K is of the same order as at atmospheric pressure. It seems that provided there is not too much blending with radiation from the outer cone, the temperature for the reaction zone measured in this way is around 5700° K at gas pressures above about 10 mm. Below this pressure, however, there appears to be a change, the temperature rising. It seems that below this pressure there is less chance of loss of energy by collision, and the departure from the theoretical flame temperature, which is very marked even at atmospheric pressure, becomes even greater. At the lowest pressure the curve is not so straight, tending to indicate a lower temperature at points involving the lower rotational energy levels, and a higher temperature for points of high K'. This indicates some departure from a Maxwell-Boltzmann distribution.

Variation with mixture composition. The rotational temperature at the centre of the flame has been determined for rich and weak mixtures, as well as the stoichiometric mixture:

	stoichiometric		rich		weak	
	$C_2H_2 + 2\frac{1}{2}O_2$		$1\frac{1}{2}C_2H_2 + 2\frac{1}{2}O_2$		$C_2H_2 + 3\frac{3}{4}O_2$	
	pressure	temp. (° K)	pressure	temp. (° K)	pressure	temp. (° K)
21 mm. burner	5·5	6200	6	6800	7	5850
34 mm. burner	3	6800	3	8600	3·3	6200

The rotational temperature is highest for mixtures containing excess acetylene. The total strength of the OH radiation does not vary rapidly with mixture composition, but is rather greater for the weak mixtures (O_2 excess).

Flames with air and nitrous oxide. The acetylene-air flame has been examined at a pressure of 25 mm. Hg using a 34 mm. burner. In this case the slope of the curve is far from constant, indicating a departure from the Maxwell-Boltzmann distribution. This is shown in figure 3. The points are rather scattered because this flame is not very luminous and it was necessary to use the smaller spectrograph with a consequent loss in accuracy. However, it is clear from the measurement of a number of plates that the curve can be regarded approximately as consisting of two straight lines, the points for spectrum lines arising from levels of low rotational energy falling on a line corresponding to a temperature of about 2850° K, while the points involving levels of high rotational energy give about 5700° K. This latter value is about the same as for an ordinary Bunsen or for an oxy-acetylene flame at this pressure.

The flame of acetylene burning with nitrous oxide has a similar characteristic, the curve being again separable into two roughly straight portions, the slopes of which give temperatures of about 2840 and 3960° K. It may also be noted that while an oxy-hydrogen flame has a rotational temperature for the (0, 0) band of OH of about the theoretical value for the low-pressure flame (2330° K), the introduction of a trace (less than 1 %) of acetylene has a marked effect. The intensity of the OH bands is considerably increased and the curve for the rotational temperature again becomes complex, the part for levels of low rotational energy remaining at a slope corresponding roughly to the theoretical temperature, while that for high rotational levels gives a very high rotational temperature, 8700° K at 7 mm., and even higher at lower pressure.

This suggests that in these flames the excitation may be by two processes. Molecules excited as a direct result of the flame reactions have a very high rotational temperature, which causes a high population for the highly excited rotational energy levels, while excitation to the lower rotational levels is mainly due to ordinary thermal processes. Points based on the higher rotational levels then give an effective temperature for the molecules excited by the reaction processes, while the points from the lower rotational levels give the temperature of the thermally excited molecules, which is usually near the theoretical flame temperature.

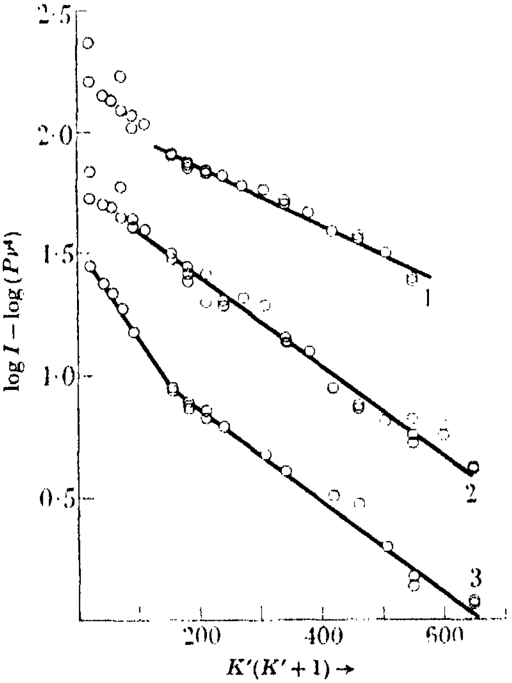


FIGURE 3. Curve 1: C_2H_2-O_2 stoichiometric at 52 mm. burner. $T = 8750°$ K. Curve 2: C_2H_2-O_2 stoichiometric at 13 mm. burner. $T = 5700$ °K. Curve 3: C_2H_2-air at 34 mm. burner. $T = 5700°$ K at high K' and 2850° at low K'.

ABSORPTION MEASUREMENTS ON OH

For the observation of the OH band in absorption it is necessary to use a spectrograph of high resolving power, such that the instrumental line width is at least comparable with the width of the absorption lines of OH. This width is about 0·3 cm.$^{-1}$, while that of the large quartz Littrow-type instrument used appeared to be about 0·7 cm.$^{-1}$. This resolving power was sufficient to enable observation of the OH absorption, but it was not possible directly to determine the concentration of OH by measuring $\int \alpha l d\nu$, where α = absorption coefficient, l = length of absorbing layer and ν = frequency. It was, however, possible to obtain an approximate value by neglecting small errors, since the absorption was weak. For a photographic plate the curve of density against log intensity is straight over a limited range. Thus, writing $I = I_0 e^{-\alpha l}$, $I - I_0$ will be proportional to αl for weak absorption. Thus if the absorption lines are measured with a photometer having a slit width greater than the line-width, it is possible to obtain an approximate value for αl. By multiplying by the width of the photometer slit, in cm.$^{-1}$, we obtain a value for $\int \alpha l d\nu$. By comparing these values of $\int \alpha l d\nu$ with those obtained by Oldenberg & Rieke (1938) who used very high dispersion to measure the OH bands in the dissociation of water vapour at about 1400° C, we can examine the concentration of OH through the flame.

For approximately stoichiometric mixtures of acetylene with oxygen it was not possible to measure the absorption by OH in the flame because the excitation temperature for the OH band exceeded that of any available background source. By using a carbon arc it was, however, possible to obtain satisfactory absorption with acetylene-air mixtures and with acetylene-oxygen mixtures containing a fair excess of oxygen.

With a burner of 21 mm. diameter the partial pressure of OH came out about 0·7 mm. both with a C_2H_2-air flame (at 30 mm. pressure) and with a weak oxy-acetylene flame (pressure 4 mm.). Somewhat similar values were obtained with a burner of 13 mm. diameter. The variation of the partial pressure of OH through the reaction zone can be seen from figures 4 and 5 in which αl is plotted against height above the top of the burner. The approximate position of the visible reaction zone is indicated, this being determined with the aid of a calibrated telescope.

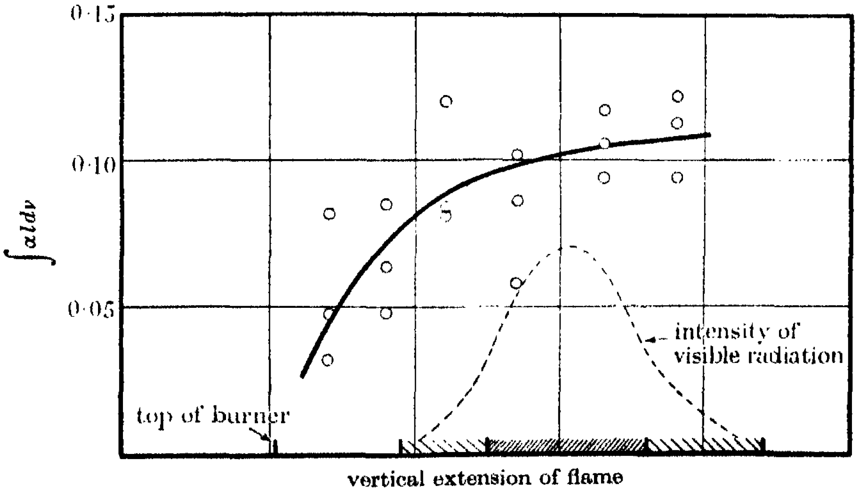


FIGURE 4. OH *concentration in a* C_2H_2-O_2 *flame measured with the lines* Q_1 ($4\frac{1}{2}$), Q_1 ($5\frac{1}{2}$), Q_1 ($6\frac{1}{2}$). Pressure 4 mm. Hg. Burner diameter 21 mm. C_2H_2 1·1 cm.^{3}/sec.; O_2 4·2 cm.^{3}/sec.

The partial pressure of OH above the reaction zone is only slightly higher than that to be expected for equilibrium at the theoretical flame temperature, but in front of the visible reaction zone, where the temperature measured with a thermocouple is about 1000° C, the concentration is far higher than for thermal equilibrium. This result may be slightly influenced by the flame overhanging the burner a little instead of being quite flat, but the main result is clear and indicates either abnormal production of OH in this region by pre-flame reactions or else diffusion of OH from the flame back against the gas stream.*

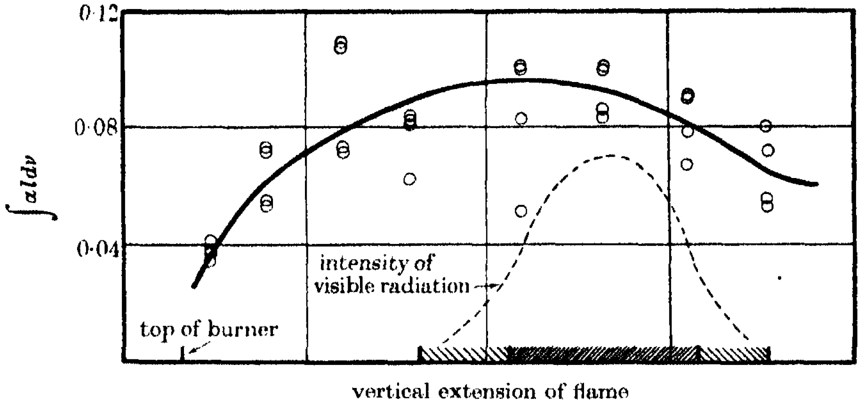


FIGURE 5. OH concentration in a C_2H_2-air flame measured with the lines Q_1 (3½), Q_1 (4½), Q_1 (5½), Q_1 (6½). Pressure 66 mm. Hg. Burner diameter 13 mm. C_2H_2 0·75 cm.^{3}/sec.; air 10·0 cm.^{3}/sec.

It has not been possible to determine the rotational temperature of the OH radicals by comparison of the strengths of the lines in absorption, but it seems unlikely that the rotational temperature measured in this way would be very different from the theoretical flame temperature, as it is in emission. In the pre-flame region where the absorption, though weak, is stronger than would be expected, the intensity maximum of the lines of the Q branch lies near the 5th or 6th Q line, which gives a temperature of around 1000° K.

THEORETICAL CALCULATIONS OF TEMPERATURE AND COMPOSITION

In order to compare the observed temperatures with the theoretical temperatures in these flames, so as to obtain some idea of the extent of departure from thermal equilibrium, the theoretical temperatures have been calculated for various mixtures. The concentrations of the various molecules and free atoms and radicals have also been calculated in some cases. These calculations have been made in the manner described by Damköhler & Edse (1943). They assume no loss of heat from the flame either by radiation or to the burner walls, and take account of the following equilibria

$$H_2O \rightleftharpoons H_2 + \tfrac{1}{2}O_2, \quad O_2 \rightleftharpoons O + O, \quad CO_2 \rightleftharpoons CO + \tfrac{1}{2}O_2.$$
$$H_2O \rightleftharpoons OH + \tfrac{1}{2}H_2, \quad H_2 \rightleftharpoons H + H,$$

It is assumed that there is no solid carbon formation, and that all fuel molecules are burnt, without formation of partially oxidized compounds such as alcohols or aldehydes. For weak, stoichiometric and slightly rich mixtures these assumptions are not likely to lead to error. With very rich mixtures some departure from the calculated values may occur through failure of these assumptions, although no carbon formation occurred under the conditions of the flames discussed here.

---

* Some preliminary calculations indicate that the high concentration is probably due to diffusion; the distance which a molecule may diffuse is of the same order as the flame thickness.

The following are the calculated temperatures of oxy-acetylene flames at atmospheric pressure and at a hundredth atmosphere,

mixture	1 atm. (°K)	$\frac{1}{100}$ atm. (°K)
$C_2H_2 + 1\frac{1}{2}O_2$	3400	2760
$C_2H_2 + 2\frac{1}{2}O_2$	3320	2710
$C_2H_2 + 5O_2$	3080	2570

The acetylene-air flame has a theoretical temperature of 2520° K at atmospheric pressure, and this temperature will only fall relatively little at the pressures used in these investigations, as the flames with air do not burn at such low pressure as with oxygen and the dissociation to CO and free atoms is less important in the cooler flame.

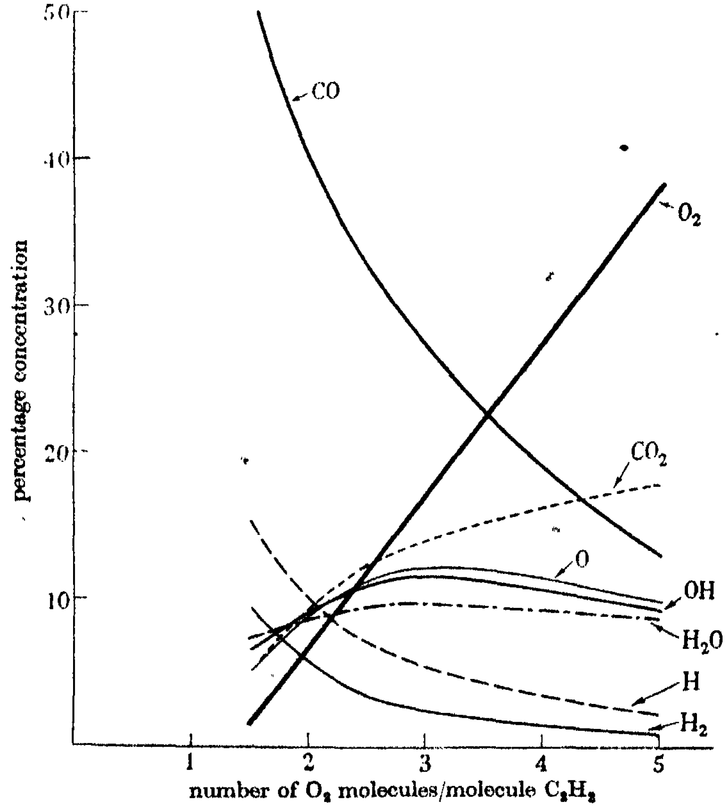


FIGURE 6. Calculated composition of a C_2H_2-O_2 flame at 1 atm.

Figures 6 and 7 show the calculated percentage composition for oxy-acetylene flames at atmospheric pressure and at one-hundredth of an atmosphere for comparison. In these very hot flames the dissociation to free atoms and of CO_2 to CO and oxygen is very considerable, but the influence of reduced pressure in increasing this dissociation is largely offset by the lower temperature. Thus the concentration of atomic oxygen rises from 12 % at 1 atm. to 14½ % at $\frac{1}{100}$ atm., that of atomic hydrogen rises from 7 to 13 % for the stoichiometric mixture, while that of OH shows a slight fall from 11½ to 9½ %. The proportion of CO_2 falls appreciably, owing to the greater dissociation to CO and oxygen at the lower pressure, and this is the main reason for the lower flame temperature.

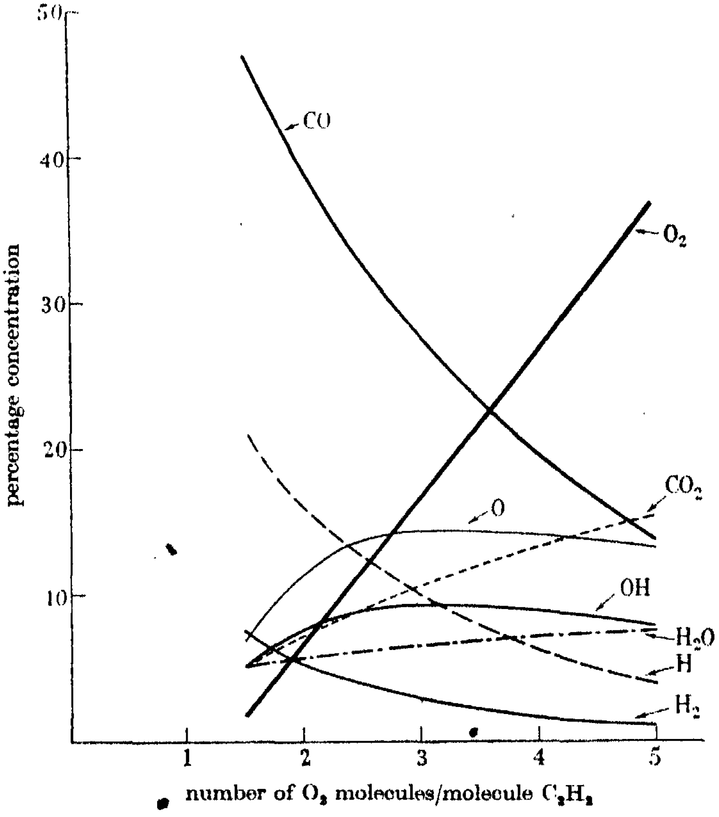


FIGURE 7. Calculated composition of a C_2H_2-O_2 flame at $\frac{1}{100}$ atm.

DISCUSSION OF RESULTS

It is quite clear from the measurements that there is a very considerable departure from thermal equilibrium in the reaction zone. The spectra of these low-pressure flames are closely similar to those of the corresponding flames at atmospheric pressure. The thickness of the zone at low pressure is about 200 times the mean free path of the gas molecules ($\lambda \sim 0{\cdot}1$ mm. at 2 mm. pressure and 2500° K); this ratio of the thickness of the reaction zone to the mean free path is about the same as at 1 atm. The number of collisions made by a molecule as it passes through the reaction zone remains at about 10^4, and it may be especially noted that the flame velocity shows hardly any change between 1 atm. and these low pressures; for the oxy-acetylene flame it is constant at about 900 cm./sec. Thus it seems most likely that for these hot flames the reaction processes are not so very different at low pressure and at atmospheric pressure. Slow combustion and ignition processes are, of course, more sensitive to changes of pressure and to surface effects. The chief difference between the hot flames at normal and low pressure is probably the smaller tendency to carbon formation, because the low pressure does not favour rapid polymerization. However, on the whole it seems that while results from low-pressure flames may not be quantitatively applicable to combustion processes at atmospheric pressure, they should serve as valuable pointers to the types of process occurring.

All these spectroscopic methods of measuring the temperature appear to give values which are much too high. This might appear to indicate that the flame contains some active species, such as for example an abnormally high concentration of fast electrons, which cause strong excitation of all substances in the flame. However, this is not supported by the observation that all the Fe lines do not reverse at the same temperature, or the fact that the rotational temperatures are in general higher than the excitation temperatures. Thus in the low-pressure flame of acetylene with air we have,

Theoretical flame temperature	2400° K
Excitation temperature by Fe lines	3500° K
Excitation temperature of OH	$<$ 3800° K
Rotational temperature excited OH	5700° K

On the other hand, the temperature of the unexcited OH radicals (observed by absorption) probably does not differ much from the theoretical equilibrium value. It seems fairly certain that in these flames with acetylene the excitation of the OH is due to chemiluminescence. This is shown by the great strength of the OH radiation from these flames; it is around 500 times greater than from an oxy-hydrogen flame which has a comparable temperature and OH concentration. Also the OH radiation is vastly stronger in the reaction zone than just above, where the radiation is no doubt mainly thermal in origin.

At pressures in excess of 10 mm. Hg the rotational temperature from the OH emission seems to be always around 5700° K (C_2H_2-air 5700° K at 25 mm. and 5400° K at 1 atm.; C_2H_2-O_2 5700° K at 13 mm.; Bunsen flame 5700° K), but at lower pressure there is some change occurring, and the temperature has a higher value. There are probably four main effects to be considered, (i) deactivation of electronically excited OH by collision, (ii) deactivation by radiation, (iii) transfer of rotational energy from one OH radical to another, and (iv) conversion of rotational energy into some other form by collision. At 10 mm. pressure and at the theoretical flame temperature the time between collisions, if we assume a normal collision diameter for excited OH, is around 10^{-7} sec., rising to 5×10^{-7} at 2 mm. pressure. We have no independent knowledge of the probability of deactivation by collision but it is known from quenching of fluorescence that in general collision processes may compete with loss of energy by radiation at pressures above a few mm., and in the case of excited OH there seems the additional chance of chemical reaction. It is known from the work of Oldenberg & Rieke (1938) that the average radiative life of excited OH is about 4×10^{-6} sec., which is rather longer than for most electronically excited molecules. Thus, if it has a normal collision diameter, an excited OH radical may make about 60 collisions during its radiative life at 10 mm. pressure, but only around 8 at the lowest pressure (1·5 mm.) used in these investigations. The emitting molecules have a very high rotational temperature while the unexcited molecules do not seem to have such a high temperature, so that there does not appear to be maintenance of equilibrium by rapid transfer of rotational energy between excited and unexcited OH. On the other hand the rotational temperature does change as we pass through the reaction zone and is markedly influenced by pressure at very low pressures; thus there is evidence that some, but not complete, conversion of rotational energy to some other form does occur during the life of the excited radicals.

It did not, at first, appear easy to reconcile these varied observations and to give a complete theory of the processes connected with the activation and deactivation of the OH radicals. However, the following hypotheses seem to give a satisfactory and co-ordinated explanation of the phenomena so far observed. It may be assumed that excited OH radicals are formed in two ways. (*a*) By thermal excitation of OH radicals; excited radicals of this type will then have a fairly normal rotational energy distribution and are responsible for the OH emission from above the reaction zone,* and in some cases for the major part of the radiation from levels of low rotational energy in the reaction zone; they thus account for the fairly normal temperatures sometimes obtained by using points at low K' on the rotational temperature curve. (*b*) By chemiluminescence; the OH radicals are assumed to be formed in an excited electronic state, perhaps by reaction of some organic molecule, and as thus formed have a very high rotational energy, above 9000° K, and have initially an approximately Maxwell-Boltzmann distribution of rotational energy. If these molecules were to lose rotational energy in large amounts at single collisions then the Maxwell-Boltzmann distribution would be destroyed, which does not appear to be the case. Therefore, we assume that the conversion of rotational energy into other forms occurs only by a unit or so at each collision so that after a number of collisions the effective rotational temperature may fall but the departure from the Maxwell-Boltzmann distribution will not be appreciable. The fundamental conclusion is now that at pressures above about 10 mm. the life of the excited OH is determined primarily by collision deactivation and that this occurs, on the average, after a moderate number of collisions, probably about 40 if excited OH radicals have a normal collision diameter. During these 40 or so collisions the radicals will have lost a part of their high rotational energy so that the emission spectrum from the reaction zone gives a sort of average rotational temperature between the initial high value and the value to which it is lowered by collisions; in practice this value appears to be around 5700° K, varying a little perhaps with gas composition. At lower pressure, however, deactivation of OH by radiation becomes relatively more important and begins to compete with collision deactivation as the factor determining the life of excited OH. Thus as we go to lower pressure the radicals make fewer collisions before losing their electronic excitation, and therefore their average rotational temperature is nearer to the high initial value of at least 9000° K.

---

* In previous work by one of us (Wolfhard 1939) the OH radiation from the inter-conal gases was found to give a rotational temperature about 200° K too high, but this value is now known to be in error by about this amount because of inadequate allowance for change of plate sensitivity and of aperture with wave-length.

Taking this as a basic explanation of the main results we may now speculate on the more detailed nature of the collisions responsible for the deactivation of electronically excited OH and the conversion of rotational energy to other forms. We may make use of five main observations. (i) When a trace of acetylene is added to an oxy-hydrogen flame at about 7 mm. pressure the rotational temperature has the very high value of 8700° K compared with about 6000° K for a pure oxy-acetylene flame at that pressure. (ii) In stoichiometric mixtures the rotational temperature falls as we pass through the reaction zone, being highest at the base of the flame. (iii) In the flame with N_2O the rotational temperature, although still abnormally high, is much lower than with O_2 or air, being about 3960° K. (iv) At very low pressure rich mixtures of acetylene with oxygen give a higher temperature than weak mixtures. (v) The limiting temperature above 10 mm. pressure with air and O_2 is not very different, being around 5700° K.

For (i), the chief difference between the pure oxy-acetylene flame, and the flame of oxy-hydrogen containing a trace of C_2H_2 (which gives the higher temperature) is the absence of appreciable quantities of CO and CO_2 from the latter; now, at the same pressure, a higher temperature could be caused by either a *more* efficient collision deactivation of *electronically* excited radicals (so that there is less time for loss of rotational energy by other collisions), or a *less* efficient conversion of *rotational* energy to other forms by collision; since absence of a particular constituent cannot account for a more efficient deactivation, the role of CO or CO_2 must be in their conversion of rotational energy. We thus conclude that the most important collisions for conversion of rotational energy of OH to other forms of energy are those with either CO or CO_2 molecules, probably the latter. Observations (ii) and (iii), the higher temperature at the base of the flame and the lower temperature with N_2O, are most easily accounted for by assuming that molecules of O_2 are mostly responsible for removal of electronically excited OH; rapid removal of excited OH gives a high temperature, while slow removal allows more time for the rotational energy to fall; in the flame with nitrous oxide there is little O_2 present, so the temperature is low (this explanation assumes that the excited OH is formed by the same reaction in the flame with N_2O as with O_2); in flames with oxygen the concentration of O_2 is highest at the base of the flame, where it causes efficient removal of excited OH and a high rotational temperature, while higher in the flame most of the O_2 is consumed, so there is less efficient removal of OH and more time for loss of rotational energy. For the explanation of (iv), the higher temperature with rich mixtures, we may note that the active molecules for collisions, O_2 and CO_2, tend to increase together for weak mixtures and to decrease together for rich mixtures; thus for weak mixtures both deactivation of electronically excited molecules and removal of rotational energy will be favoured, so that the 'transition point' for change over from collision to radiative deactivation as the dominant factor will lie at lower gas pressure than for rich mixtures; thus at equivalent pressures in the neighbourhood of the transition point the rich mixtures will give higher rotational temperatures than the weak ones. Observation (v), the nearly equal temperatures with air or oxygen, is most simply explained by assuming that nitrogen molecules are relatively inactive both for removal of electronic and rotational energy.

The high excitation temperatures in the flames as determined with Fe lines does not seem to be directly related to the excitation of OH. It is possible that collisions with excited OH radicals, leading to excitation of Fe, may raise the average effective temperature of the Fe atoms. It is also likely that there is some incomplete equipartition of energy in other forms in the reaction zone of the flame. In its passage right through the reaction zone a molecule makes of the order 10^4 collisions, but the number of collisions made by molecules which are actually formed in some part of the reaction zone may be appreciably smaller. It is unlikely that complete equipartition of vibrational energy is maintained or that recombination of free atoms can keep pace with changing conditions. Such factors may clearly be able to give temperatures appreciably above or below the theoretical values for special forms of excitation.

The present observations have established some facts about the degree of departure from equipartition of energy in the reaction zone, especially for the excited OH radicals. It is hoped to extend this type of work to other flames and to apply it to other radicals such as C_2 and CH. Quantitative measurements of light yield for various flames at various pressures are also likely to improve our understanding of flame reactions.

In conclusion we wish to express our sincere thanks to Sir Alfred Egerton for his interest in these researches, and one of us (A.G.G.) is indebted to the Royal Society for the award of a Warren Research Fellowship.

REFERENCES

Damköhler, G. & Edse, R. 1943 *Z. Elektrochem.* **49**, 178.
Earls, L. T. 1935 *Phys. Rev.* **48**, 423.
Gaydon, A. G. 1947 *Internat. Congr. Pure and App. Chem., London.*
Gaydon, A. G. 1948 *Spectroscopy and combustion theory*, 2nd ed.
Gaydon, A. G. & Wolfhard, H. G. 1948 *Trans. Faraday Soc.* (in the Press).
Klaukens, H. & Wolfhard, H. G. 1948 *Proc. Roy. Soc.* A, **193**, 512.
Kondratjew, V. & Ziskin, M. 1937 *Acta Physiochim. U.R.S.S.* **6**, 307.
Lewis, B. & von Elbe, G. 1943 *J. Chem. Phys.* **11**, 75.
Oldenberg, O. 1934 *Phys. Rev.* **46**, 210.
Oldenberg, O. & Rieke, F. F. 1938 *J. Chem. Phys.* **6**, 439.
Wolfhard, H. G. 1939 *Z. Phys.* **112**, 107.
Wolfhard, H. G. 1943 *Z. tech. Phys.* **9**, 206.

'Diffuse scattering' of the Fermi electrons in monovalent metals in relation to their electrical resistivities

By A. B. Bhatia, D.Phil. and Sir K. S. Krishnan, F.R.S.

(*Received* 27 *November* 1947)

The main step in the calculation of the electrical resistivities of monovalent metals, in which the conduction electrons are almost completely degenerate, is the calculation of the relaxation time τ of the electrons at the Fermi surface, which in these metals is a sphere, and is well inside the first Brillouin zone. Since the wave-length λ, and hence the group velocity v, of the Fermi electrons is known, the calculation of τ means also the calculation of the mean free path $l = v\tau$ of these electrons. Now the finite mean free path of these electrons arises from the scattering—particularly the large-angle scattering—of these electrons in their passage through the crystal, by the thermally agitated atoms. Hence a detailed knowledge of the scattering coefficient of the crystal for the Fermi electrons, incident and scattered along different directions in the crystal, will enable us to calculate τ or l.

Now the scattering coefficient depends on two factors.

1. The atom form factor for scattering, which in monovalent metals may be taken to be isotropic, i.e. independent of the direction of incidence or of scattering separately, but *dependent on the angle of scattering* ϕ *between them, and on* λ. (*Extensive measurements* are available on the scattering of slow electrons by the rare gases, which give us information regarding the atom form factors for the scattering of the Fermi electrons in the corresponding alkali metals, and the variation of these factors with ϕ.)

2. The structure factor of the crystal, which, besides being a function of λ, will vary, even in a cubic crystal, with the direction of incidence and of scattering, but will, however, be independent of the nature of the waves, i.e. independent of whether they are X-rays, or electron or neutron waves. (The 'diffuse scattering' of *X-rays* of *long* wave-lengths in crystals has been studied in great detail, both theoretically and experimentally, from which one can calculate the structure factors of the monovalent metals for their respective Fermi wave-lengths, for different directions of incidence and of scattering in the crystals.)

Using these data for the atom form factor and for the structure factor of the crystal, the mean free path of the Fermi electrons is calculated in detail in the present paper for different directions of incidence, for one typical monovalent metal, namely sodium crystal. The free path l is given by

$$1/l = \Psi\nu^2 kT\beta\sigma,$$

where ν is the number of atoms per unit volume, σ is the cross-section of the atom for total scattering in all directions, β is the compressibility, and Ψ is a numerical factor which varies from a maximum of about 2·2 for incidence along [110] to a minimum of about 0·9 for incidence along [100], its average value being close to the minimum, and nearly unity.

With Ψ actually unity, the right-hand side of the above expression for $1/l$ can be seen to be just *the Einstein-Smoluchowski expression for the attenuation coefficient of a liquid medium for long waves*: which shows that in sodium, and presumably in the other monovalent metals also, the mean free path of the Fermi electrons may be taken roughly as the reciprocal of the attenuation coefficient of the crystal due to scattering, and the scattering may be regarded as due almost wholly to the local thermal fluctuations in density, and the Fermi wave-length as long enough for the Einstein-Smoluchowski formula for density-scattering to be applicable.

1. INTRODUCTION

In an electronic conductor, e.g. a metal crystal, it is only the electrons near the Fermi surface that can make effective collisions and thus take part in conduction. Hence a detailed knowledge of the coefficient of scattering along different directions in the crystal, of the Fermi electrons, will be helpful in our understanding its electrical properties. In general the wave-lengths of these electrons are of the order of the interatomic distance. Now the scattering coefficient of a crystal depends both on the atom form factor, i.e. the cross-section of the atom for scattering, and on the structure factor of the crystal. For waves much shorter than the interatomic distance, or much longer, both these factors are easily evaluated, but for intermediate wave-lengths, such as are involved in electrical conduction, the evaluation is more difficult, but has been done by earlier workers. The atom form factor for scattering of electrons of such wave-lengths has been studied theoretically by Faxén & Holtsmark (1927), and experimentally by Ramsauer & Kollath (1932), and others. The technique for the calculation of the structure factor of the crystal for scattering of these wave-lengths is also available, since the structure factor for 'diffuse scattering' of X-rays has been studied extensively, both theoretically and experimentally, and the structure factor of a crystal for the scattering of electrons—or of any other waves—is the same as for the scattering of X-rays of the same wave-length. (In the usual discussions on diffuse scattering of X-rays, however, attention is directed mainly to the neighbourhood of the Bragg spots, whereas in the scattering of the Fermi waves we shall be concerned mainly with the region between the central spot and the immediately surrounding Bragg spots.)

In *good* metals like the alkali or the noble metals, i.e. metals of the first group in the Periodic Table, in which the Fermi surface is well within the first Brillouin zone, and is a sphere, the calculation of the scattering coefficient for the Fermi electrons becomes much simpler, since all the Fermi electrons have the same wave-length, and this wave-length is too long to give any Bragg reflexion; in the alkali metals, which are body-centred cubic, this wave-length is 14 % longer, and in the noble metals, which are face-centred cubic, 11 % longer, than that necessary to give the first Bragg reflexion in the backward direction.

Detailed calculations are given in the present paper of the structure factor of sodium crystal for scattering of wave-lengths too long to give any Bragg reflexion. Some rough estimates can be made of the atom form factor for scattering of the Fermi electrons by the atoms in the metal. These data are then utilized to calculate the relaxation time of the Fermi electrons in the metal, and thence its electrical resistivity.

In a previous paper (Krishnan & Bhatia 1944) it was shown from certain general considerations, that in the alkali and the noble metals the mean free path of the Fermi electrons may be taken roughly to be equal to the reciprocal of the attenuation coefficient of the metal for the passage of these electrons, the attenuation being due to scattering by the thermally agitated atoms in the metals; and that for the Fermi wave-length the scattering may be taken to be due predominantly to the local thermal fluctuations in density and the attenuation coefficient to be given practically by the Einstein-Smoluchowski formula. The detailed calculations made in the present paper for sodium crystal show that these results are good approximations for this metal, and presumably for other monovalent metals too. It is needless to add that they simplify considerably the calculation of electrical resistivities.

2. Electrical resistivity in relation to the scattering of the Fermi electrons

As is well known the specific resistance ρ of a *good* metal or alloy is given by the expression

$$\rho = \frac{m}{n\nu e^2}\frac{1}{\tau} \tag{1}$$

$$= \frac{h}{n\nu e^2 \lambda}\frac{1}{l}, \tag{2}$$

where τ is the relaxation time of the Fermi electrons, of wave-length $\lambda = (8\pi/3n\nu)^{\frac{1}{3}}$ and of group velocity $v = h/m\lambda$, $l = v\tau$ is their mean free path, ν is the number of atoms per unit volume, and n is the number of conduction electrons per atom, and the other letters have their usual significance (see Mott & Jones 1936; Fröhlich 1936). All the quantities in (1) and (2) are known except τ (or l), and the main problem in calculating the electrical resistivities of good metals or alloys is therefore one of calculating τ (or l) in terms of the scattering of the Fermi electrons by the thermally agitated atoms.

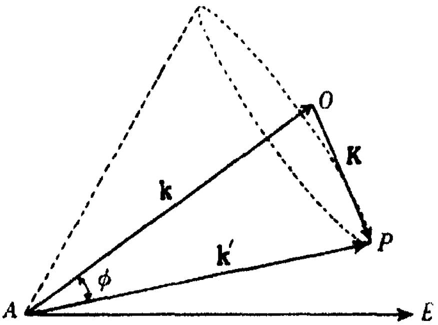


Figure 1

Let AE (see figure 1) be the direction of the applied electric field. Consider electrons at the Fermi surface moving along any given direction AO, and scattered along AP, the angle of scattering being ϕ. The wave-vectors $\mathbf{k}$ and $\mathbf{k}'$ ($k' \approx k$) of the incident and the scattered electron-waves respectively, are as marked in the figure. $\mathbf{K}$ is the wave-vector of the elastic wave that produces the scattering, where $K \approx 2k \sin \frac{1}{2}\phi$. Consider now the quantity $\Delta p_E/p_E = \Delta k_E/k_E$, where $p = hk$ is the momentum of the incident Fermi electron, and $\Delta p = h\Delta k$ is the change in its momentum due to scattering, and the subscripts E denote the components along AE of these quantities. $\mathbf{\Delta k}$ will evidently be equal to $\mathbf{K}$.

Let $S_{kk'}$ be the fraction of the Fermi electrons moving along direction $\mathbf{k}$ which is scattered per unit volume of the crystal, per unit solid angle,* along direction $\mathbf{k}'$. The reciprocal of the mean free path appearing in (2) will then be given by

$$\frac{1}{l} = \frac{1}{v\tau} = \frac{1}{4\pi}\iint S_{kk'}\frac{\Delta k_E}{k_E}\,d\omega_k\,d\omega_{k'}, \tag{3}$$

---

* The probability that a Fermi electron travelling along $\mathbf{k}$ is scattered per unit time per unit solid angle along $\mathbf{k}'$ is given by

$$P_{kk'} = S_{kk'}v,$$

where v is the velocity of the electron.

where $d\omega_k$ and $d\omega_{k'}$ are elements of solid angles about the directions **k** and **k**′ respectively, and both the integrations extend over the whole of the 4π solid angle. It will be seen from (3) that $1/l$ may also be regarded as 4π times the average value of $S_{kk'}\Delta k_E/k_E$, the average being taken over all directions of scattering **k**′, and all directions of incidence **k** of the Fermi electrons.*

Even in a cubic crystal, because of its elastic anisotropy, the scattering coefficient $S_{kk'}$ will be a function of **k** and **k**′. Even so, with a detailed knowledge of $S_{kk'}$ as a function of **k** and **k**′, which as we shall show in the present paper is not difficult to obtain, the integration in (3) can be done numerically.

In the usual discussions on the electrical resistivities of these metals, it is assumed for convenience that $S_{kk'}$ is a function of the angle of scattering ϕ between **k** and **k**′, and is independent of **k** and **k**′ separately. This assumption will hold when the crystal is elastically isotropic, and when further the Fermi wave-length is much longer than is necessary to give any Bragg reflexion, i.e. $\lambda \gg 2d \sin \frac{1}{2}\phi$, where d is the lattice spacing corresponding to the first Bragg reflexion. Neither of these conditions, however, is satisfied in the alkali or the noble metals, since many of them are elastically anisotropic, and λ, as we have seen, is only slightly greater than $2d$. Holding for the present to the assumption that $S_{kk'}$ is a function of ϕ alone, which we can denote by S_ϕ, we may evaluate (3) in a simple manner, since we can now keep one of the directions **k**, **k**′, say **k** fixed, and find the average value of $S_\phi \Delta k_E/k_E$ over all directions of the other, namely **k**′. We can do this averaging in two steps. First we keep ϕ, and hence S_ϕ, constant, and average $\Delta k_E/k_E$ over all *azimuths* of scattering; the average is just $K \sin \frac{1}{2}\phi/k$, or $(1-\cos\phi)$, which gives for $1/l$ the value

$$\frac{1}{l} = \frac{1}{v\tau} = \int_0^\pi S_\phi(1-\cos\phi)\, 2\pi \sin\phi\, d\phi, \tag{4}$$

which is independent of the direction of incidence **k** in the crystal.

Though we do not adopt the approximation that $S_{kk'}$ is a function of ϕ alone, and is independent of **k** or **k**′ separately, we shall have occasion to use expression (4), since even in elastically anisotropic crystals like the alkali metals there are regions in reciprocal space in which for certain directions of incidence the scattering coefficient is roughly a function of ϕ alone.

3. DIFFUSE SCATTERING AS DUE TO THERMAL AGITATION

The influence of the thermal agitations of the atoms in a crystal on the intensities of the Bragg reflexions and of the 'diffuse scattering' along other directions, has been discussed among the earlier investigators by Darwin (1914) and by Debye (1914). Let R be the structure factor of the crystal, defined as usual by

$$R = \frac{S}{\nu s},$$

---

* In the derivation of expression (1) it is implicit that $1/\tau$ is the same for all the Fermi electrons, i.e. for all directions of incidence since in the metals that we are considering all the Fermi electrons have the same wave-length. When $1/\tau$ varies with the direction of incidence, one may as an approximation, use the average value of $1/\tau$ taken over all directions of the Fermi electrons, as we have done in expression (3).

where s is the scattering coefficient of an atom in the crystal per unit solid angle, or the atom form factor for intensity, all the three quantities R, S and s referring to the given wave-length and to *the same directions of incidence and of observation*. In the *monatomic* metals that we are considering, s may be regarded as practically isotropic, i.e. a function of only ϕ and λ in the combination $\sin \frac{1}{2}\phi/\lambda$.

The major effect of the thermal spread of the atomic centres about their mean positions will be to make the atom form factor at temperature T, namely s_T, fall off more rapidly with increase of $\sin \frac{1}{2}\phi/\lambda$ than the corresponding form factor s_0 of the non-vibrating atom, according to the well-known expression

$$\left.\begin{aligned} s_T &= s_0 e^{-2M}, \\ M &= 8\pi^2\overline{\xi^2}(\sin \tfrac{1}{2}\phi/\lambda)^2, \end{aligned}\right\} \tag{5}$$

where $\overline{\xi^2}$ is the mean square of the amplitude of the thermal oscillations taken along the external bisector of the angle between the directions of incidence and of scattering, and is given at room temperature, which in these metals is sufficiently high in comparison with the Debye temperature to make the thermal energy proportional to T, by

$$4\pi^2\nu_v^2 m\overline{\xi^2} \approx kT,$$

where ν_v is the Einstein vibration frequency and m the mass of the atom. The result is a rapid fall of intensity of the Bragg reflexions particularly of higher orders, with rise of temperature. This is observed, and the fall is roughly in accordance with (5).

The question naturally arises as to what happens to the energy represented by the diminution of the intensities of the Bragg reflexions consequent on increase of temperature. Obviously it should appear as diffuse scattering along other directions, and besides, if a sufficient number of higher order Bragg reflexions is included in the range $0 < \phi \leqslant \pi$, we should have for any given direction of incidence

$$\nu\int s_0\,d\omega - \nu\int s_T\,d\omega = \int S\,d\omega, \tag{6}$$

where the integrations extend over the whole of the 4π solid angle of scattering.

If the only effect of the thermal agitations may be regarded as a static swelling up or inflation of the scattering atoms, as is implied in the above treatment, then, in addition to the integrated value of S over all directions being equal to the difference between the integrated values of νs_0 and of νs_T over these directions, as given by (6), S will be equal to $\nu(s_0 - s_T)$ along *every* direction. Though some extensive measurements by Jauncey and his collaborators (1935) apparently lend support to this conclusion, it is now realized, both from theoretical considerations, and from scattering experiments made with a narrow parallel pencil of incident monochromatic X-rays, which verify the theory, that such a detailed balancing between $\nu(s_0 - s_T)$ and S does not occur.

4. CALCULATION OF DIFFUSE SCATTERING

A detailed calculation of the directional distribution of intensity of the diffuse scattering of X-rays in crystals due to thermal agitation has been made by Faxén (1923) and by Waller (1923, 1925, 1928), and more recently, and from a different point of view, by Raman (1941). We may mention immediately that in crystals in which *all* the normal modes of vibration are of the acoustic type, e.g. those of the alkali and the noble metals, the Faxén-Waller treatment and the treatment by Raman give identical results. (The criterion adopted by Raman for distinguishing between the 'acoustic', and the 'optical', normal modes of vibration of a crystal, is not, however, the same as the criterion adopted here.)

The Faxén-Waller treatment has been simplified, or made more rigorous, by Zachariasen, Born, Weigel, and Jahn,* among others. We shall use here the results as presented by Jahn (1942*a*) since they are in a form directly applicable to our problem.

Consider the reciprocal lattice of a given cubic crystal and the first Brillouin zones surrounding each of the points of the reciprocal lattice, including (000). These zones will be filled by a quasi-continuous and uniform distribution of cells, each cell defining a wave-vector **K** associated with the elastic standing waves that represent the normal modes of vibration of the crystal. Each K-cell will correspond to *three* normal modes of vibration of the crystal answering respectively to the three directions of vibration or displacement—and the three frequencies—associated with the particular value of **K**.

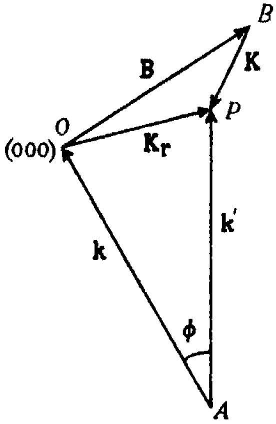


FIGURE 2

In the alkali metals, which are body-centred cubic, the first Brillouin zone surrounding each point of the reciprocal lattice is the rhombic dodecahedron bounded by faces of the type {110}. The distance from the reciprocal point forming the centre of the dodecahedron to the faces of the dodecahedron will be $1/2d_{110}$, where d_{110} is the grating spacing of the (110) planes in the crystal. In the noble metals, which are face-centred cubic, the first Brillouin zone is the orthotetrakaidekahedron, or the 14-hedron, bounded by {111} and {200} faces, i.e. the octahedron {111} truncated by the cubic faces {200}. In both the alkali and the noble metals *these first zones can accommodate just one K-cell per atom, i.e. all the normal modes of vibration of the crystal*

---

* For a good account of the subject see symposia of papers on the subject in *Proc. Ind. Acad. Sci.* **14**, 1941, and in *Proc. Roy. Soc.* A, **179**, 1942, and the reports by Mrs Lonsdale and by Max Born in *Reports on Progress in Physics*, Physical Society, London, 1942–43.

will be of the acoustic type, and further the first zones surrounding the various reciprocal lattice points will be similar and similarly oriented, and these zones will fill up the whole of the reciprocal space without overlapping. (As is well-known, both the rhombic dodecahedron and the orthotetrakaidekahedron described above are 'space-fillers'.)

Consider now a narrow pencil of monochromatic electron waves of wave-number $k = 1/\lambda$ incident along AO (see figure 2), where O is the origin (000) of the reciprocal lattice, and consider the radiations scattered along AP, where $AP = k' \approx k$, and ϕ is the angle between AO and AP or the angle of scattering as before. Let P—which, following Laval, we may call the 'pole of diffusion'—lie in the zone associated with the reciprocal lattice point B. Let $\mathbf{B}$ denote the vector OB, and $\mathbf{K}$ the elastic wave-vector from B to P. The vector $\mathbf{K}_r$ defined by

$$\mathbf{K}_r = \mathbf{B} + \mathbf{K}$$

will bisect externally the angle between the direction of incidence $\mathbf{k}(AO)$ and the direction of observation $\mathbf{k}'(AP)$. The structure factor R of the crystal for the given direction of incidence and of scattering will be given, according to Jahn, by*

$$R = e^{-2M} \frac{\nu k T}{\rho} \left\{ \left(\frac{\mathbf{K}_r \cdot \mathbf{e}_1}{N_1}\right)^2 + \left(\frac{\mathbf{K}_r \cdot \mathbf{e}_2}{N_2}\right)^2 + \left(\frac{\mathbf{K}_r \cdot \mathbf{e}_3}{N_3}\right)^2 \right\}, \tag{7}$$

where ρ is the density of the medium, N_1, N_2, N_3 are the three frequencies and $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ the three unit polarization vectors, i.e. unit vectors along the directions of displacement of the atoms, associated with the particular elastic wave-vector $\mathbf{K}$.

Following again Jahn, we may rewrite (7) in the form

$$\begin{aligned} R = e^{-2M} \frac{\nu k T}{\rho} \Bigg[& B^2 \sum_{i=1}^{3} \frac{(LA_i + MB_i + NC_i)^2}{N_i^2} \\ & + 2BK \sum_{i=1}^{3} \frac{(LA_i + MB_i + NC_i)(lA_i + mB_i + nC_i)}{N_i^2} \\ & + K^2 \sum_{i=1}^{3} \frac{(lA_i + mB_i + nC_i)^2}{N_i^2} \Bigg], \end{aligned} \tag{8}$$

where L, M, N are the direction cosines of $\mathbf{B}$, l, m, n those of $\mathbf{K}$, and A_i, B_i, C_i those of the polarization vectors $\mathbf{e}_i$, $i = 1, 2, 3$.

In the special case when the pole of diffusion P lies in the central Brillouin zone, i.e. in the zone associated with (000), the first two terms in (8) vanish, since for this zone $B = 0$ (and $\mathbf{K}_r \equiv \mathbf{K}$), and we are left with the third term alone. The expression for R then reduces to

$$R = e^{-2M} \frac{\nu k T}{\rho} \sum_{i=1}^{3} \frac{(lA_i + mB_i + nC_i)^2}{V_i^2}, \tag{9}$$

where $V_i = N_i/K$ is the phase velocity of the elastic wave.

---

* Except for a constant of proportionality which is left undetermined in Jahn's expression, but is easily evaluated when we regard scattering as reflexion from the appropriate thermal elastic waves in the crystal (Krishnan & Bhatia 1948), and is included in (7).

5. SCATTERING REGARDED AS REFLEXION FROM APPROPRIATE THERMAL ELASTIC WAVES

In a previous paper, on the scattering of light-waves, i.e. of very long waves, in homogeneous media (Krishnan & Bhatia 1948), we showed that the scattering coefficient S may also be regarded as equal to the maximum value of the coefficient of reflexion from the two sets of progressive elastic waves (the two progressing in opposite directions) whose wave-normals and wave-lengths satisfy the Bragg condition for reflexion, either set consisting of three progressive waves corresponding to the three directions of displacement, the energy of each of the waves, *filling the whole of the medium*, being that of a Planck oscillator of the same frequency, or kT under the present assumptions. The expression for R comes out on this basis as (Krishnan & Bhatia 1948)

$$R = \frac{\nu kT}{\rho} \sum_{i=1}^{3} \frac{\cos^2 \omega_i}{V_i^2}, \tag{10}$$

where ω_i is the angle between $\mathbf{K}$ and $\mathbf{e}_i$.

Expression (10), deduced here for long waves, can be seen to be identical with the Faxén-Waller expression (9), except for the factor e^{-2M}, which for long waves is practically unity.

For shorter wave-lengths also, i.e. when the pole of diffusion P lies outside the central Brillouin zone, R can be calculated in terms of reflexions from the elastic waves, and the expression thus obtained is found to be identical with the corresponding expression (7) of Faxén-Waller. We demonstrate this equivalence below.

Consider any one set of atomic reflecting planes (hkl) in the crystal, which, under suitable conditions, can give rise to a Bragg reflexion, and let the spacing of these layers be d_{hkl}. This regular layered structure may be regarded as the limiting case of a plane condensational acoustic wave with its wave-normal perpendicular to the planes, its wave-length equal to d_{hkl}, and *its velocity of propagation zero.* Actually since the distribution of matter in these layers is not sinusoidal, but is concentrated near the plane of the atomic centres, we shall have to take into account the upper partials also, corresponding to the spacings $1/2$, $1/3$, etc. times d_{hkl}. These shorter grating spacings may be regarded as giving rise to Bragg reflexions of the second, third and higher orders, corresponding to the reciprocal lattice points $2(hkl)$, $3(hkl)$, etc., and since these reciprocal lattice points are considered separately, we shall confine ourselves here to the fundamental spacing d_{hkl} corresponding to the point (hkl) of the reciprocal lattice.

Consider now one of the elastic waves representing a normal mode of vibration of the crystal, whose wave-vector is $\mathbf{K}$, and consider also a regular stratification of the crystal of spacing d_{hkl}. Because of this stratification the above wave may also be regarded as an elastic wave whose wave vector $\mathbf{K}_r$ is given by

$$\mathbf{K}_r = \overrightarrow{\left(\frac{1}{\mathbf{d}_{hkl}}\right)} + \mathbf{K}.$$

We shall refer to such waves obtained by the superposition of elastic waves on the stratified layers of the scattering medium, as *resultant* elastic waves, as distinguished from the *primary* elastic waves. Whereas the wave numbers of the latter, namely $\mathbf{K}$, are restricted in our crystals to the limits of the first Brillouin zone, the wave numbers $\mathbf{K}_r$ of the resultant elastic waves may have very large values, $K_r \approx n/d_{hkl}$.

The frequency of the resultant elastic wave, the displacement of the atoms under it, and the energy per unit volume associated with it, will all be the same as for the primary elastic wave producing it, but the velocity of the resultant wave will naturally be smaller than that of the primary wave in the ratio

$$V_r/V = K/K_r.$$

When the primary wave is a long one, such that $K \ll K_r$ the velocity of the resultant wave can become quite small.

These resultant waves can reflect the incident X-rays in the same manner as the primary elastic waves, and thus account for scattering along directions corresponding to positions of P outside the central Brillouin zone, i.e. along directions for which there can be no direct reflexion from any of the primary waves.

It will be seen from relation (10) that when the total energy in the medium associated with the elastic wave remains fixed, namely at kT, the intensity of reflexion from the elastic wave, which determines the scattering along the direction concerned, is inversely proportional to the square of its velocity, and hence the intensity of reflexion from the resultant wave, because of its low velocity, will be larger than that of the reflexion from the primary wave by a factor $(K_r/K)^2$, which, as K becomes small, or as P approaches the reciprocal point corresponding to the Bragg reflexion from the particular set of reflecting planes considered, will become very large.

The intensity of scattering along these directions will also be conditioned by the following.

(1) Just as in reflexion by a primary elastic wave, i.e. when P lies in the central Brillouin zone, it is the components of the three displacements along the wave-normal $\mathbf{K}$ of the reflecting elastic wave that determine the intensity of scattering, so also in the case of reflexion by the secondary elastic wave referred to above, i.e. when P lies outside the central Brillouin zone, it is the components of the displacements along $\mathbf{K}_r$ that determine the intensity, but the absolute directions of these displacements in the crystal will be determined by the direction of propagation (lmn) of the primary wave alone, according to relation (12) to be given in the next section.

(2) For the short wave-lengths λ that we are considering now, the temperature factor e^{-2M} will deviate considerably from unity, and will have to be retained.

These considerations obviously lead to an expression for R identical with expression (7) given above, of Faxén & Waller.

For these short waves we should remember that the atom form factor s_0 will vary markedly with ϕ unlike for long waves.

6. REMARKS ON THE CALCULATION OF τ FOR SOME SPECIAL DIRECTIONS OF ELECTRON WAVES

Consider the locus of the pole of diffusion P in reciprocal space when the direction of incidence is kept fixed and the direction of scattering is varied over the whole of the 4π solid angle. The locus will evidently be a sphere passing through the origin, of diameter $2k$, the diameter through the origin being along the direction of incidence. Following the usual practice we shall call this the 'sphere of propagation'. Since equal areas on the sphere correspond to equal solid angles of scattering, the calculation of $1/l$ for the given direction of incidence is equivalent to integrating

$$\frac{1}{k^2} S_{kk'} \frac{\Delta k_E}{k_E} dA,$$

where dA is an element of area on the surface of the sphere of propagation, over the whole of the sphere.

We saw in § 2 that when S is a function of ϕ alone, the average value of $S_\phi(\Delta k_E/k_E)$ over all azimuths of scattering corresponding to a given angle of scattering ϕ, is just equal to $S_\phi(1-\cos\phi)$. Even when the crystal is cubic like ours, $S_{kk'}$ is not a function of ϕ alone. If, however, the incidence is along [100], [110] or [111], which are axes of tetragonal, digonal and trigonal symmetry respectively, and we consider all the *crystallographically equivalent* directions ϕ, θ_i corresponding to the same angle of scattering ϕ and to equivalent azimuthal angles θ_i, then $S_{kk'} = S_{\phi\theta_i}$ will be the same for all these directions of scattering, and the average value of $\Delta k_E/k_E$ over these directions will again be $(1-\cos\phi)$. Hence for the above special directions of incidence in a cubic crystal, the expression for $1/l$ may be written in the form

$$\frac{1}{l} = \int_{\phi=0}^{\pi} \int_{\theta=0}^{2\pi} S_{kk'}(1-\cos\phi)\sin\phi\, d\phi\, d\theta. \tag{11}$$

In other words the weightage factor that takes into account the relatively large contribution to $1/l$ from large angle scattering, is equal to $(1-\cos\phi)$, and hence the same as when $S_{kk'}$ is a function of ϕ alone.

7. SPHERE OF PROPAGATION WHOLLY INSIDE THE FIRST BRILLOUIN ZONE

Before proceeding to calculate $1/l$ for the alkali and the noble metals, for their respective Fermi wave-lengths, for which the sphere of propagation extends into the second Brillouin zone also, it will be useful to calculate its value for the special case when λ is large, i.e. when the sphere of propagation is *well* within the first Brillouin zone. This case is simple since then for every point P on the sphere the velocity of propagation, and the directions of the displacements, are independent of the *wave-number* K of the elastic wave, though of course they will vary with the *direction* of $\mathbf{K}$; and V_t, and A_t, B_t, C_t, occurring in (9), can then be expressed in terms of the static elastic constants c_{11}, c_{12} and c_{44} of the cubic crystal, by the relations

$$\left.\begin{aligned}\{c_{11}l^2 + c_{44}(m^2+n^2) - \rho V_i^2\}A_i + (c_{12}+c_{44})(lmB_i + nlC_i) &= 0,\\ \{c_{11}m^2 + c_{44}(l^2+n^2) - \rho V_i^2\}B_i + (c_{12}+c_{44})(lmA_i + mnC_i) &= 0,\\ \{c_{11}n^2 + c_{44}(l^2+m^2) - \rho V_i^2\}C_i + (c_{12}+c_{44})(lnA_i + mnB_i) &= 0.\end{aligned}\right\} \tag{12}$$

In view of these relations, expression (9) reduces, as Jahn has shown, to

$$R = e^{-2M}\nu kT \frac{c_{44}^2 + 2c_{44}\epsilon(l^2m^2+m^2n^2+n^2l^2) + 3\epsilon^2 l^2m^2n^2}{c_{11}c_{44}^2 + c_{44}(c_{11}+c_{12})\epsilon(l^2m^2+m^2n^2+n^2l^2) + (c_{11}+2c_{12}+c_{44})\epsilon^2 l^2m^2n^2} \tag{13}$$

$$\equiv e^{-2M}\nu kT \,.\, F_{lmn}, \text{ say,} \tag{14}$$

where ϵ is the elastic anisotropy of the crystal, given by

$$\epsilon = c_{11} - c_{12} - 2c_{44}.$$

Moreover in this region e^{-2M} is practically unity, and s_ϕ practically independent of ϕ, and equal to $\sigma/4\pi$, where σ is the cross-section of the atom for *total scattering in all directions*, and is given by

$$\sigma = \int_0^\pi s_\phi \, 2\pi \sin\phi \, d\phi. \tag{15}$$

Consider incidence of long electron waves along one of the principal directions [100], [110] or [111] of the cubic crystal. Expression (11) for $1/l$ reduces under the above conditions to the simple form

$$1/l = \nu^2 kT(\sigma/4\pi) J, \tag{16}$$

where
$$J = \int_{\phi=0}^{\pi} \int_{\theta=0}^{2\pi} F_{lmn}(1-\cos\phi)\sin\phi \, d\phi \, d\theta. \tag{17}$$

When the crystal is elastically isotropic, i.e. when $\epsilon = 0$, F becomes independent of lmn, and equal to $1/c_{11}$. In that case expression (17) holds whatever may be the direction of incidence, and the factor $(1-\cos\phi)$ appearing in (17) intended to give the proper weightage for scattering at large angles, over scattering at small angles, becomes quite ineffective ultimately, and the expression for $1/l$ merely reduces to

$$1/l = \mu = \nu^2 kT\sigma/c_{11}, \tag{18}$$

where μ is the total scattering in all directions, per unit volume of the crystal, or the attenuation coefficient of the crystal for the passage of electron waves along the direction considered, the attenuation being due to scattering, and the coefficient μ being defined by

$$I_d = I_0 e^{-\mu d},$$

where I_d is the intensity after traversing a distance d in the crystal.

If further $c_{44} = 0$, c_{11} becomes equal to c_{12} and to $1/\beta$, where β is the compressibility of the crystal, and the expression for the attenuation coefficient due to scattering then reduces to the Einstein-Smoluchowski value for a liquid (Krishnan & Bhatia 1948), namely

$$1/l = \mu = \nu^2 kT\beta.\sigma \tag{19}$$

as it should.

Coming back to the elastically anisotropic cubic crystal, and considering first incidence of long electron waves along $[\bar{1}00]$, we may express F_{lmn} defined by (13) and (14), in the form*

$$F_{lmn} = \frac{a + b(l^2m^2 + m^2n^2 + n^2l^2) + cl^2m^2n^2}{a' + b'(l^2m^2 + m^2n^2 + n^2l^2) + c'l^2m^2n^2}$$
$$= \frac{A + B\sin^2 2\theta}{A' + B'\sin^2 2\theta}, \tag{20}$$

where
$$A = a + \tfrac{1}{4}b\sin^2\phi, \quad B = \tfrac{1}{4}\cos^4\tfrac{1}{2}\phi(b + c\sin^2\tfrac{1}{2}\phi),$$

and A' and B' have similar values in terms of a' and b'.

---

* We should remember here that lmn are the direction cosines of the elastic wave-vector **K** which bisects externally the angle between **k** and **k'**, and when **k** is along $[\bar{1}00]$, is related to the angle ϕ, and the azimuth θ, of scattering by the relations

$$l = \sin\tfrac{1}{2}\phi, \quad m = \cos\tfrac{1}{2}\phi\cos\theta, \quad n = -\cos\tfrac{1}{2}\phi\sin\theta.$$

In view of expression (20) for F in which A, B, A', B' are all independent of θ, the integration in (17) with reference to θ can be done readily. Doing so we obtain

$$J = \int_0^{\pi} Y(1-\cos\phi)\sin\phi\, d\phi, \tag{21}$$

where

$$Y = \int_0^{2\pi} F_{lmn}\, d\theta$$

$$= 2\pi\left\{\frac{B}{B'} + \frac{AB'-A'B}{B'(A'+B')}\sqrt{\frac{A'+B'}{A'}}\right\}. \tag{22}$$

Changing the variable in (21) by putting $f = 1-\cos\phi$ we obtain

$$J = \int_0^{2} Yf\, df. \tag{23}$$

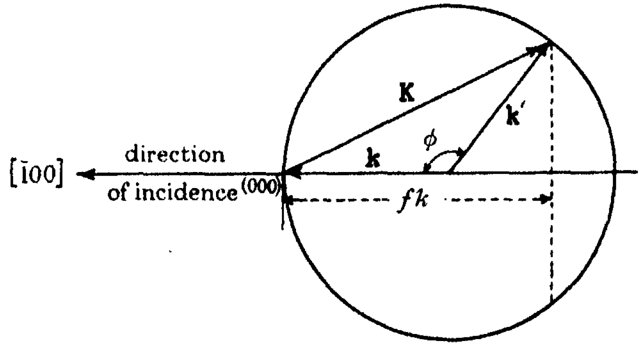


FIGURE 3

Obviously fk represents the perpendicular distance from (000) of the plane perpendicular to the direction of incidence which by its intersection of the sphere of propagation defines ϕ (see figure 3).

The integration in (23) can be done numerically, since we can calculate readily the values of Y corresponding to different values of f in the range $0 \leqslant f \leqslant 2$.

The elastic constants of sodium crystal have been determined by Quimby & Siegel (1938) at different low temperatures, which on extrapolation, give for room temperature (see Lonsdale & Smith 1940)

$$c_{11} = 0{\cdot}52, \quad c_{12} = 0{\cdot}40, \quad c_{44} = 0{\cdot}41,$$

all of them in 10^{11} dynes/cm.^{2}. Or

$$c_{12} = 0{\cdot}77c_{11}, \quad c_{44} = 0{\cdot}79c_{11}, \quad \epsilon = -1{\cdot}35c_{11}. \tag{24}$$

Using these data, we have calculated the values of Y for different values of f in the range $0 \leqslant f \leqslant 2$ at intervals of 0·2. The values are entered in table 1.

TABLE 1

f	0 forward direction	0·2	0·4	0·6	0·8	1·0	1·2	1·4	1·6	1·8	2·0 backward direction
$Y.c_{11}/2\pi$	0·84	0·77	0·69	0·60	0·53	0·50	0·59	0·74	0·86	0·95	1·0

It will be seen from the expression for B' that it becomes zero when $\sin^2 \frac{1}{2}\phi = -b'/c'$, or when $f = 0{\cdot}62$. Since B' appears as a multiplying factor in the denominator of both the terms in expression (22) for Y, it appears at first sight that Y may become infinite at $f = 0{\cdot}62$. But actually on making B' small, and expanding the two terms and taking them together and then making B' tend to zero, the Y *vs* f curve is found to decrease smoothly at this point.

B' becomes again zero at $\phi = \pi$, i.e. at $f = 2$; but it can be readily seen that both the terms in (22) are separately finite at this point since B' occurs in the combination B/B', and B/B' remains finite at $\phi = \pi$.

Using the values given in table 1, and Simpson's rule for numerical integration, we obtain

$$J = 2\pi \times 1{\cdot}52/c_{11},$$

and hence for incidence of long electron waves along [100]

$$1/l = 0{\cdot}76\nu^2 kT\sigma/c_{11}, \tag{25}$$

as against

$$1/l = \nu^2 kT\sigma/c_{11}$$

for the elastically isotropic crystal (see (18)).

The calculation of $1/l$ for other directions of incidence in the crystal, even [110] or [111], becomes laborious. But one can see in a general way that for long waves $1/l$ will be roughly of this magnitude for all directions of incidence.

We may also note here that because of the $1-\cos\phi$ factor, the bulk of the contribution to $1/l$ comes from large angle scatterings, the small angle ones contributing very little to it.

8. VARIATION OF R WITH THE WAVE-LENGTH λ

In the region of reciprocal space close to (000), corresponding to $\lambda \gg 2d \sin \frac{1}{2}\phi$, the velocities V_i and the directions of displacements $\mathbf{e}_i$ defined by A_i, B_i, C_i, and hence also R, will be independent of the magnitude of the elastic wave-vector $\mathbf{K}$ (but will of course depend on the direction of $\mathbf{K}$). This will not be true, however, when the pole of diffusion P approaches the boundary of the Brillouin zone. One has a rough idea how at least the velocity varies as one approaches the boundary, which may be taken to indicate in a general way the variation of the displacements also. Calculations have been made by Kellerman (1940) of the phase velocities V_i of the elastic waves in rock salt crystal. Though the velocities decrease as one approaches the boundary, the total drop, even for the [111] direction, for which it is a maximum, is only 25 %, and much of this drop is confined to the close neighbourhood of the boundary. Calculations have also been made for crystals of tungsten (Fine 1939), and of KCl (Blackman 1937), of the density of distribution of the normal modes of vibration of the crystal, on the *frequency* scale, i.e. $p(N)$ as a function of the frequency N, where $p(N)\,dN$ gives the number of normal modes of vibration whose frequencies lie between N and $N+dN$. Now the corresponding distribution on the K scale, i.e. $P(K)$ as function of K, where $P(K)\,dK$ gives the number of normal modes of vibration between K and $K+dK$, is known to be uniform over the whole of the

Brillouin zone. A comparison of the distributions on the two scales, using the simple relations*

$$N/K = V \quad \text{and} \quad dN/dK = G, \tag{26}$$

where V is the phase-velocity, and G the group-velocity of the elastic waves, gives a rough estimate of the change in the phase-velocity as we approach the boundary of the zone. These estimates also show that the total drop in the phase-velocity as we approach the boundary is a fraction, of the order of 1/4, most of the change again occurring in the neighbourhood of the boundary. (The group-velocity G will of course diminish much more rapidly than V in this region, and tend to zero at the boundary of the zone.) The result will be an excess of scattering in the close neighbourhood of the boundary of the Brillouin zone, which, however, being highly localized, will contribute little to the total scattering taken over all directions.

TABLE 2. (SODIUM CRYSTAL)

K_r along [100]

$K_r d_{110}$	$<0{\cdot}707\dagger = 1/\sqrt{2}$	0·75	0·80	0·877‡
$R\dfrac{c_{11}}{e^{-2M}\,\nu kT}$	1·0	1·31	1·70	2·73

TABLE 3. (SODIUM CRYSTAL)

K_r along [110]

$K_r d_{110}$	$<0{\cdot}5\dagger$	0·55	0·60	0·70	0·80	0·85	0·877‡
$R\dfrac{c_{11}}{e^{-2M}\,\nu kT}$	0·60	0·89	1·34	3·12	9·55	19·2	30·5

TABLE 4. (COPPER CRYSTAL)

K_r along [100]

$K_r d_{111}$	$<0{\cdot}58\dagger = 1/\sqrt{3}$	0·60	0·70	0·80	0·90‡
$R\dfrac{c_{11}}{e^{-2M}\,\nu kT}$	1·0	1·17	2·36	5·08	12·6

TABLE 5. (COPPER CRYSTAL)

K_r along [111]

$K_r d_{111}$	$<0{\cdot}5\dagger$	0·55	0·60	0·70	0·80	0·85	0·90‡
$R\dfrac{c_{11}}{e^{-2M}\,\nu kT}$	0·71	1·06	1·60	3·72	11·4	22·8	58·8

When we move outside the first Brillouin zone, the value of R, which is now determined by $\mathbf{K}_r$, will naturally vary not only with the direction of $\mathbf{K}_r$ but also with its magnitude, as can be seen from (7). In order to get an idea of the variation of R

---

* In the special case when the medium is non-crystalline and $P(K) = 3 \times 4\pi K^2$, we obtain

$$p(N)\,dN = 3 \times 4\pi N^2 dN/(V^3 G).$$

† These values of K_r correspond to the upper limit of the first Brillouin zone. For values of K_r greater or smaller than this limiting value, *but not close to it*, R can be calculated on the basis of formula (8), and the values entered in the tables are those calculated on the basis. When K_r is close to this limiting value, the value of R given by (8) will be too low, as we mentioned earlier in this section.

‡ These values of K_r correspond to backward scattering ($\phi = \pi$) of electrons having the Fermi wave-length.

in this region, we have calculated for sodium the variation of R with $\sin \frac{1}{2}\phi/\lambda$ for two directions of $\mathbf{K}_r$, namely [100] and [110]. The special advantage of these directions is that for all positions of P on these lines $\mathbf{K}_r$ can be obtained in a unique way, and the directions of $\mathbf{B}$, $\mathbf{K}$ and $\mathbf{K}_r$ are identical (neglecting the sense of the direction) so that R will continue to be a function of $\sin \frac{1}{2}\phi/\lambda$, as in the central region.

The values of R for different positions of P along [100] and [110] for sodium crystal, and along [100] and [111] for copper crystal, are entered in tables 2 to 5. The distances from the origin are expressed in terms of $1/d_{110}$ for sodium, and in terms of $1/d_{111}$ for copper, i.e. in terms of the distance in reciprocal space of the first Bragg spot from (000).

9. THE ATOM FORM FACTOR *s* AND ITS VARIATION WITH ϕ

We have seen that for the Fermi wave-length the sphere of propagation extends much beyond the first Brillouin zone, and that for the regions on the sphere remote from (000) R may vary rapidly. Though this will be the major factor affecting S, two other factors also affect it considerably.

(1) The factor e^{-2M}, which in the region near (000) is practically unity, will now be appreciably less, and will diminish with increase of ϕ. For sodium crystal at room temperature (300° K), and for its Fermi wave-length

$$\lambda = 2{\cdot}28 d_{110} = 6{\cdot}93\,\text{A}\,(3{\cdot}12\,\text{eV}), \quad e^{-2M} = e^{-0{\cdot}27 \sin^2 \frac{1}{2}\phi}$$

decreases by about 24 % as we pass from the forward to the backward direction.*

TABLE 6

ϕ (degrees)	s_ϕ (A^2)	ϕ (degrees)	s_ϕ (A^2)
15	0·220	105·5	0·148
28	0·214	121	0·131
43	0·218	137	0·120
59	0·220	152	0·117
74·5	0·207	167	0·112
90	0·173		

(2) The atom form factor s, which in the region near (000) is practically independent of ϕ will not be so now. Knowing, however, the Fermi wave-length λ, and the electronic distribution in the atoms of these metals, it is possible theoretically to calculate the variation of s with ϕ. The calculation is rather laborious. But we can get the data concerning the variation of s with ϕ required for our present purpose from the extensive measurements made by Ramsauer & Kollath (1932) on the scattering coefficients S_ϕ of rare gases for different angles of scattering ϕ and for different wave-lengths λ. The scattering medium in their experiments being gaseous, these data give directly the atom form factors s_ϕ, since $s_\phi = S_\phi/\nu$. The data for neon gas for 3·12 eV, i.e. for electrons having the same wave-length as the Fermi electrons in sodium metal, are given in table 6.

---

* The value $2M = 0{\cdot}27 \sin^2 \frac{1}{2}\phi/\lambda^2$ given here is on the basis of the Debye temperature of sodium being 150° K.

It will be seen from table 6 that s_ϕ is practically independent of ϕ up to about $\phi = 75°$, after which it drops down, reaching about half the value at about $\phi = 125°$, and remains afterwards practically constant at this value. The general variation of s_ϕ with ϕ observed here holds not only for 3·12 eV, but also for electrons having energies in the neighbourhood of this value, and for other rare gases too. We may, therefore, take this result to be applicable to the scattering atoms in the alkali metals also, and as a rough approximation take s_ϕ to be constant and equal to s_1 over the whole of the range $0 \leqslant \phi \leqslant \phi_0$, and again constant and equal to $s_2 = \frac{1}{2}s_1$ over the remaining angles of scattering $\phi_0 < \phi \leqslant \pi$, where ϕ_0 is about 100°.

10. CALCULATION OF τ FOR THE FERMI ELECTRONS IN SODIUM CRYSTAL

We now proceed to calculate $1/l$ for one typical crystal, namely sodium, and for its Fermi wave-length of 6·93 A, when the direction of incidence is along $[\bar{1}\bar{1}0]$, i.e. to evaluate the integral

$$\frac{1}{l} = \int_{\phi=0}^{\pi} \int_{\theta=0}^{2\pi} S_{kk'}(1-\cos\phi)\sin\phi\, d\phi\, d\theta$$

over the whole of the corresponding sphere of propagation. *The value of $1/l$ will be a maximum for this direction of incidence in the alkali metals.* For the purpose of the integration we split the surface of the sphere into two parts, namely that included in the central zone, and that outside respectively. The critical angle of scattering ϕ_0 defining the transition from the former to the latter region is about 98°, corresponding to (see § 7) $f_0 = 1 - \cos 98° = 1·14$, and in view of the remarks made at the end of the previous section we may to a first approximation take s_ϕ to be constant and equal to s_1 over the whole of the former part, $0 \leqslant \phi \leqslant 98°$, and again constant and equal to $s_2 = s_1/2$ over the latter part $98° < \phi \leqslant \pi$, the average value of s over the whole of the sphere being given by

$$\bar{s} = \sigma/(4\pi) = \tfrac{1}{2}(1·14s_1 + 0·86s_2) = 0·79s_1. \tag{27}$$

Let us denote by $(1/l)_1$ the contribution to $1/l$ from the part of the sphere included in the first Brillouin zone, corresponding to $0 \leqslant \phi \leqslant 98°$, and by $(1/l)_2$ the contribution from the part of the sphere that lies outside the first zone and corresponding to $98° < \phi \leqslant \pi$.

The calculation of $(1/l)_1$ is done immediately. Since $(1/l)_1$ is the contribution to $1/l$ from forward angles of scattering and is a small part of $1/l$, we may in view of (23) take it to be just $(f_0/2k)^2$ times the value of $1/l$ for the complete sphere of propagation, *the whole of the sphere being taken for this purpose to be well within the first Brillouin zone*, or

$$\begin{aligned}(1/l)_1 &= \nu^2 kT \,.\, 4\pi s_1(0·76/c_{11})\,(1·14k/2k)^2 \\ &= 0·3\nu^2 kT\sigma/c_{11}.\end{aligned} \tag{28}$$

The contribution to $1/l$ from the part of the sphere outside the central zone, which we have denoted by $(1/l)_2$, and which is the major contribution to $1/l$, is calculated in the following manner, remembering that in this region also s_ϕ may be taken to be independent of ϕ, but equal to s_2. The major factor that determines variation of S

in this region will be the variation of R, which will be particularly marked in the regions of the sphere facing the reciprocal point (110). Jahn (1942*b*) has discussed in the case of X-rays the variation of R for scattering by sodium crystal, for different positions of P in the neighbourhood of the reciprocal point (110), and has investigated in particular the shape of the 'isodiffusion' surfaces, i.e. the loci of P corresponding to different constant values of the coefficient R_ϕ. The portions of these isodiffusion surfaces in the neighbourhood of the line [110] and the reciprocal point (110), are found to be portions of highly oblate ellipsoids, with their short axis along [110], and hence the intersections of these surfaces with the sphere of propagation will be practically planes perpendicular to the above line and perpendicular to the plane of the paper in figure 4.

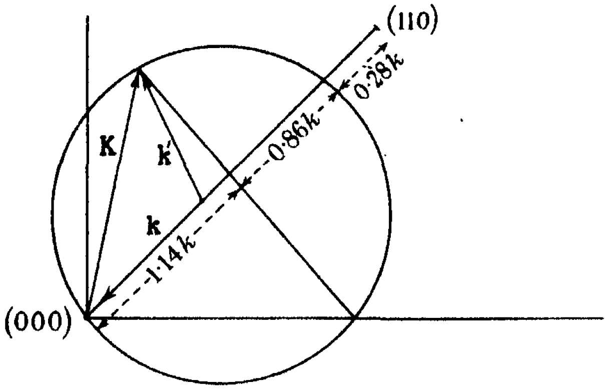


FIGURE 4

In order to give an idea how flat these surfaces are, we give in table 7 for a few of the isodiffusion ellipsoids surrounding (110), the magnitudes of the semi-axes along the three principal directions of the ellipsoid, which are also the directions of displacement $\mathbf{e}_i$, I along [110]; II along [$\bar{1}$10]; and III along [001]. The semi-axes are all expressed in terms of the Fermi wave-number $k = 1/(2{\cdot}28d_{110})$ as unit. (The shortest distance from (110) to the surface of the sphere of propagation is $0{\cdot}28k$.) The calculation of the axes is based on the data for the elastic constants of sodium crystal given in § 7.

TABLE 7. SEMI-AXES OF SOME OF THE ISODIFFUSION ELLIPSOIDS AROUND (110) AND CLOSE TO THEIR [110] AXES

I	II	III	$R\dfrac{c_{11}}{e^{-2M}\,\nu kT}$
0·28	1·23	0·48k	30·5
0·4	1·90	0·73	13·2
0·5	2·55	1·00	7·6
0·75	4·90	2·10	2·48
0·9	7·06	4·04	1·40
1·0	10·89	—	0·98

It will be seen that the II and III semi-axes are much longer than I, especially as we move out some distance from (110) and as the area of intersection of these isodiffusion surfaces with the sphere of propagation expands; and hence we may take these intersections practically as planes perpendicular to [110].

This feature of the isodiffusion surfaces makes the computation of the structure factor R_ϕ, and of the scattering coefficient S_ϕ, for the portions of the sphere of propagation outside the central Brillouin zone, and facing (110), and the numerical integration of $S_\phi(1-\cos\phi)$ over this region of the surface of the sphere, easy; since for the purpose of this integration we may divide the surface of the sphere in the neighbourhood of B into zones of equal area, by equally spaced planes perpendicular to [110]. In other words S in this region is determined wholly by the distance fk from (000) measured along [110], where $f = 1-\cos\phi$, $98^\circ < \phi \leqslant \pi$.

On this basis expression (7) for R, for the part of the sphere outside the central zone at a distance fk from (000) measured along [110] will be given by

$$R = e^{-2M}\frac{\nu kT}{\rho V_l^2}\left(\frac{K_r}{K}\right)^2$$
$$= e^{-2M}\frac{\nu kT}{\rho V_l^2}\left(\frac{f}{2{\cdot}28-f}\right)^2, \tag{29}$$

where V_l is the velocity of the longitudinal elastic wave of wave-number K travelling along [110], and is given by (see (12))

$$\rho V_l^2 = \tfrac{1}{2}(c_{11}+c_{12}+2c_{44}), \tag{30}$$

which for sodium crystal becomes $c_{11}/0{\cdot}60$. Hence we obtain

$$R = e^{-2M}\nu kT(0{\cdot}60/c_{11})f^2/(2{\cdot}28-f)^2 \tag{31}$$

and
$$\left(\frac{1}{l}\right)_2 = \int_{\phi_0}^{\pi} R\nu s_2(1-\cos\phi)\,2\pi\sin\phi\,d\phi$$
$$= \frac{\nu^2 kT\sigma}{c_{11}}\times 0{\cdot}190\int_{1{\cdot}14}^{2} e^{-2M}\frac{f^3 df}{(2{\cdot}28-f)^2}. \tag{32}$$

The integral can be evaluated numerically, and is found to be equal to about 12·0, whence we obtain

$$(1/l)_2 = 2{\cdot}3\nu^2 kT\sigma/c_{11}. \tag{33}$$

From (28) and (33) we obtain
$$1/l = (1/l)_1 + (1/l)_2$$
$$= 2{\cdot}6\nu^2 kT\sigma/c_{11}. \tag{34}$$

Though the above method of calculating $(1/l)_2$ by taking the intersections of the sphere of propagation by the isodiffusion surfaces to be *planes* is a good approximation, it can be readily seen that it is on the side of an over-estimate, since R will fall off slightly more rapidly as we move away from (110) than is implied in the above approximation. Hence we may express $1/l$ for the Fermi electrons in sodium crystal in the form

$$1/l = \Phi\,\nu^2 kT\sigma/c_{11}, \tag{35}$$

where for incidence along [110], Φ is slightly less than 2·6, which as we mentioned *is the highest value of Φ for any direction of incidence of the Fermi electrons.*

11. INCIDENCE OF FERMI ELECTRONS ALONG [100]

The calculation of Φ for other directions of incidence of the Fermi electrons is more laborious, but we can see in a general way that in a body-centred cubic crystal, just as Φ is a maximum for incidence along [110], it is a minimum for incidence along [100]. For the latter direction of incidence the portion of the sphere of propagation that is included in the first Brillouin zone varies with the azimuthal plane of scattering, and for (001) and similar azimuthal planes corresponds to $0 \leqslant f \leqslant 0{\cdot}66$. Now the structure factor R does not reach a high value in *any* part of the sphere of propagation, as it does when the incidence is along [110], the highest values of $Rc_{11}/(e^{-2M}\nu kT)$ now reached being only 2·7 (see table 2), for the backward direction of scattering, i.e. for $f = 2$, and falling down to unity as we move back to the boundary of the Brillouin zone. The above value of R would correspond to

$$S_{\text{max.}} = e^{-0{\cdot}27} \times 2{\cdot}7\nu^2 kT s_2/c_{11}$$

$$= 1{\cdot}3\frac{\nu^2 kT}{c_{11}}\frac{\sigma}{4\pi}.$$

Hence for this direction of incidence Φ should lie between 1·3 and 0·76, and a rough estimate shows it is about 1·1.

12. THE AVERAGE VALUE OF $1/l$ TAKEN OVER ALL DIRECTIONS OF INCIDENCE

Having fixed the approximate upper and lower limits for Φ, namely 2·6 and 1·1, for incidence of the Fermi electrons along [110] and [100] respectively, it now remains to determine the average value of Φ taken over all directions of incidence of these electrons. The high value of $\Phi = 2{\cdot}6$ for incidence along [110] is due to the close proximity of the extreme outer regions of the sphere of propagation to the reciprocal lattice point (110); indeed of the total value of 2·6 for Φ for this direction of incidence, about 1·7 is contributed by the small region $1{\cdot}75 < f \leqslant 2$, the rest of the sphere, namely $0 \leqslant f \leqslant 1{\cdot}75$, contributing about 0·9 only. Hence as the direction of incidence is varied even slightly away from [110], Φ will drop down considerably, and hence one can reasonably take the average value of Φ to be very close to its minimum value of 1·1, and take the mean value of $1/l$ or $1/v\tau$ required in calculating the specific resistance of sodium to be given by

$$\overline{1/l} = \overline{\Phi}\, \nu^2 kT\sigma/c_{11},$$

where $\overline{\Phi}$ is about 1·2 or 1·3.

Till now we have expressed $1/l$ for sodium in terms of its elastic constant c_{11}. We may also, for convenience, express it in terms of its compressibility β, which in a cubic crystal is given by

$$\beta = 3/(c_{11} + 2c_{12}),$$

and which in sodium is equal to $3/(2{\cdot}54c_{11}) = 1{\cdot}18/c_{11}$. The mean free path $1/l$ can then be expressed in the form

$$\overline{1/l} = \Psi\, \nu^2 kT\beta\sigma, \tag{36}$$

where Ψ is a numerical factor, which in sodium crystal will be seen to be close to unity. With $\Psi = 1$, one can easily recognize the right-hand side of equation (36) to be just the Einstein-Smoluchowski expression for the attenuation coefficient of a liquid medium for *long* waves (see Einstein (1910)).

Now in a liquid, the local inhomogeneities produced by thermal agitation, which ultimately account for the scattering, is represented almost wholly by the local fluctuations in density. In a crystalline medium, however, the local inhomogeneities which produce scattering will naturally be more complicated, and will involve, besides the compressibility β, the other elastic constants too. As a result of this, in sodium crystal, and for small values of $\sin \frac{1}{2}\phi/\lambda$, the structure factor R is considerably less than the Einstein value, almost half of it when **K** is along [110]. On the other hand, when the pole of diffusion P moves out of the first Brillouin zone, R as we have seen rises above the Einstein value, and for certain directions of incidence, and in the neighbourhood of the backward direction, becomes several times the Einstein value. But since these high values are highly localized their contribution to $1/l$ is not so large. Indeed the average value of $1/l$ for the Fermi electrons, taken over all directions of incidence and of scattering, as we have seen, is roughly just the Einstein value, *as though we might regard the mean free path l of the Fermi electrons in sodium to be just the reciprocal of the attenuation coefficient due to scattering, and the scattering to be determined almost wholly by the local thermal fluctuations in density.* This simple approximation is a result of the rough balancing of the different varying factors that determine the mean free path, R_ϕ and $1-\cos\phi$ increasing, e^{-2M} and s_ϕ decreasing, as we pass from the forward to the backward directions of scattering.

Had λ been considerably shorter, i.e. had the Fermi surface been much closer to the boundary of the Brillouin zone, owing to the rapid increase of R as one approaches the boundary, and the consequent increase of the scattering coefficient S, the free path and the relaxation time would be much shorter. The result would be a correspondingly enhanced specific resistance.

In the usual treatment of the electrical resistivities of metals (see Mott & Jones 1936) for convenience in calculation, the medium is regarded as elastically isotropic, and the longitudinal thermal elastic waves alone as contributing to the scattering, which is equivalent to regarding the scattering as produced by the fluctuations in density due to thermal agitation. The detailed calculations made in the present paper show that in sodium metal, this is a good approximation, and it is presumably so in other monovalent metals too.

Now the Fermi wave-length λ, as we have seen, is not much longer than necessary to give the first Bragg reflexion in the backward direction, i.e. not much longer than $2d$. Even so, it was shown in an earlier paper (Krishnan & Bhatia 1944) from certain general considerations, that it should be regarded as long enough for the Einstein formula—which is derived explicitly on the basis of λ's being very long in comparison with the interatomic distance—to be applicable. The present calculations show that the intensity of scattering, particularly along the backward direction will deviate considerably from the Einstein value, but the total scattering in all directions, i.e. the attenuation coefficient, remains roughly at this value.

13. FERMI SURFACE CLOSE TO THE BOUNDARY OF THE BRILLOUIN ZONE

We have already emphasized that if the electronic wave-length were to be reduced considerably below the Fermi value, the total scattering, and hence the specific resistance, would increase very rapidly. This would be attributed in the usual theory of electrical conduction in metals to the approach of the Fermi surface to the boundary of the Brillouin zone, and the consequent enhanced influence of the ionic lattice on the freedom of movement of the electrons. The effect would be regarded as equivalent to either an increase in the effective mass of the electron near the boundary of the zone, or a decrease in the effective number of conduction electrons, whereas from the point of view adopted in the present paper the effect is an enhanced scattering and a corresponding reduction in the mean free path of the Fermi electrons. All the three alternatives are of course equivalent, and correspond to an enhanced specific resistance. But the viewpoint adopted here, which regards the enhancement of the resistance when the Fermi surface approaches any of the reciprocal points as due to enhanced scattering of the Fermi electrons, has certain advantages, especially in dealing with the electrical resistivities of liquid metals (Krishnan & Bhatia 1945). We postpone detailed consideration of this aspect to a separate paper.

REFERENCES

Blackman, M. 1937 *Proc. Roy. Soc.* A, **159**, 416.
Darwin, C. G. 1914 *Phil. Mag.* **1**, 315.
Debye, P. 1914 *Ann. Phys., Lpz.*, **43**, 49.
Einstein, A. 1910 *Ann. Phys., Lpz.*, **33**, 1294.
Faxén, H. 1923 *Z. Phys.* **17**, 266.
Faxén, H. & Holtsmark, J. 1927 *Z. Phys.* **45**, 307.
Fine, P. C. 1939 *Phys. Rev.* **54**, 355.
Fröhlich, H. 1936 *Theorie der Metalle.* Berlin: Springer.
Jahn, H. A. 1942*a* *Proc. Roy. Soc.* A, **179**, 320.
Jahn, H. A. 1942*b* *Proc. Roy. Soc.* A, **180**, 476.
Jauncey, G. E. M. See Compton, A. H. & Allison, S. H. 1935 *X-rays in theory and experiment.* London: Macmillan & Co.
Kellerman, E. W. 1940 *Phil. Trans.* A, **238**, 513.
Krishnan, K. S. & Bhatia, A. B. 1944 *Proc. Nat. Acad. Sci. India*, **14**, 153.
Krishnan, K. S. & Bhatia, A. B. 1945 *Nature*, **156**, 503.
Krishnan, K. S. & Bhatia, A. B. 1948 *Proc. Roy. Soc.* A, **192**, 181.
Lonsdale, K. & Smith, H. 1940 *Nature*, **149**, 21.
Mott, N. F. & Jones, H. 1936 *Theory of properties of metals and alloys.* Oxford Univ. Press.
Quimby, S. L. & Siegel, S. 1938 *Phys. Rev.* **54**, 293.
Raman, Sir C. V. 1941 *Proc. Ind. Acad. Sci*, **13**, 1.
Ramsauer, C. & Kollath, R. 1932 *Ann. Phys., Lpz.*, **12**, 529.
Waller, I. 1923 *Z. Phys.* **17**, 398.
Waller, I. 1925 *Diss.* (*Uppsala*).
Waller, I. 1928 *Z. Phys.* **51**, 213.

A generalization of the method of separating longitudinal and transverse waves in electrodynamics

By Sonja Ashauer, *Newnham College, Cambridge*

(*Communicated by P. A. M. Dirac, F.R.S.—Received* 16 *July* 1947)

A generalization of the method of separating longitudinal and transverse waves in electrodynamics is proposed in this paper. It consists in splitting up each Fourier-component of the wave-field 4-vector with respect to two null-vectors $\mathbf{k}$ and $\mathbf{l}_k$, where $\mathbf{k}$ is the Fourier vector of propagation, and $\mathbf{l}_k$ is an arbitrary (real) function of $\mathbf{k}$ satisfying $(\mathbf{l}_k, \mathbf{k}) = 1$ and $\mathbf{l}_{-k} = -\mathbf{l}_k$. This amounts to referring each Fourier-component of the field to a *different* time-axis in the usual method of splitting up. $\mathbf{l}_k$ may also depend on the co-ordinates of the particles of the system. The longitudinal field variables are eliminated from the Hamiltonian formulation of electrodynamics by a contact transformation; the term which then replaces the longitudinal field depends on the co-ordinates of the particles, and in general also on the transverse field variables, except when $\mathbf{l}_k$ is independent of the co-ordinates of the particles. The work holds both in the classical and in the quantum theory.

For the problem of an electron moving in the field of a nucleus a particular form is chosen for the $\mathbf{l}_k$, which depends on the initial velocities of both the particles. The new longitudinal waves are eliminated, and the classical deflexion formula is derived on the assumption that the interaction with the new transverse waves can be neglected. The new method leads to half the deflexion derived by the usual method of working only with the Coulomb interaction force (and so neglecting radiation damping in a certain way), when the nuclear recoil is neglected and the deflexion angle is small. For a low-velocity electron, the effect of the new transverse waves cannot be small, but for a very high energy electron the new method may be more suitable than the usual one.

Introduction

The main problem of electrodynamics is to treat the interaction of a system of charged particles with an electromagnetic field. In order to enable the direct passage to the quantum theory, it is necessary to put classical electrodynamics into Hamiltonian form. In this form of the theory, each particle of the system is described by a suitable Hamiltonian for the particle and the field. In field theory, the mutual interaction between the charged particles is looked upon as an indirect one: each particle influences the neighbouring field and this influence spreads out to the other particles. The total field consists of a transverse and a longitudinal part, the latter occurring only in the presence of charged particles. It was first shown by Fermi (1930, 1932) that the longitudinal part can be eliminated from the Hamiltonian function and replaced by the static Coulomb interaction of the particles.

A Hamiltonian formulation of electrodynamics which takes into account the effect of radiation reaction has been given by Dirac (1939, 1943) with the help of the λ-limiting process. Each particle is then described by a Hamiltonian equation involving a new field (represented by its 4-vector potential) whose properties differ slightly from those of the Maxwell field and which is called the Wentzel field. This is a wave-field and can be split up into longitudinal and transverse waves according to the usual method in a Lorentz frame having a particular direction as time-axis. This method consists in taking the whole of the time-component of the field to be longitudinal, and separating the spatial part of the field into a rotation-free (longitudinal) part and a divergence-free (transverse) part. It is then shown that the longitudinal waves can be eliminated from the Hamiltonian equations by means of a contact transformation, and replaced by terms giving rise to the Coulomb interaction between the particles. This procedure simplifies the equations of motion considerably, as it reduces the number of field variables in them by half. The relativistic form of the theory, however, is spoilt.

There is great arbitrariness in the method of separating the longitudinal and transverse waves under the requirement that the divergence-free character (in four dimensions) of the transverse waves is preserved. In the usual method, all the Fourier-components of the field are split up with reference to the same selected direction (the time-axis of the particular Lorentz frame). In this paper a general method of separation is proposed, in which each Fourier-component is split up with reference to a different (arbitrary) direction, which may depend on other variables as well as on the Fourier-component. Under certain conditions the new longitudinal waves can be eliminated from the Hamiltonian equations by means of a contact transformation. The work is carried out here in the classical theory and is so arranged that each step can be taken over directly into the quantum theory.

The generalized splitting up of the field may be useful in cases where the usual method is not the most suitable. As an interesting application, the problem of the motion of two interacting charged particles is considered. In this case the usual method of eliminating the longitudinal field and treating the interaction between the particles and the transverse field as small gives satisfactory results for slowly moving particles. The neglect of the terms containing the transverse field variables in the Hamiltonian equations leads to the equations of motion for the particles derived from the Coulomb potential, from which the usual classical formula for the deflexion of an electron by a heavy nucleus is obtained (Darwin 1913). This involves the neglect of radiation damping in a certain way. For particles of large relative velocity, however, some alternative method of treatment of the field may be more suitable. A new way of separating the longitudinal and transverse field based on the proposed generalization is therefore introduced for this case. The neglect of the terms containing the new transverse field in the Hamiltonian equations leads to new equations for the motion of the particles. A new formula for the classical deflexion of a fast electron by a nucleus is thus derived and compared with the usual one.

1. GENERAL SEPARATION OF LONGITUDINAL AND TRANSVERSE WAVES

Relativistic notation will be used, in the metric defined by

$$g_{00} = -g_{11} = -g_{22} = -g_{33} = 1, \quad g_{\mu\nu} = 0 \quad \text{for} \quad \mu \neq \nu,$$

the velocity of light being unity. The scalar product of two 4-vectors **a**, **b** is denoted by

$$(\mathbf{a}, \mathbf{b}) \equiv a_\mu b^\mu \quad (\mu = 0, 1, 2, 3).$$

Consider a number of charged point-particles interacting with an electromagnetic field in the classical theory. Let e_i, $\mathbf{z}_i$, $\mathbf{p}_i$ be the charge, co-ordinates and momenta respectively of the ith particle, and $\mathbf{A}(\mathbf{x})$ the field at $\mathbf{x}$ (4-vector potential). The equations of motion of the dynamical variables are derived from the many-time Hamiltonian equations written in the usual way in terms of the momentum vector

$$\mathbf{p}_i - e_i \mathbf{A}(\mathbf{z}_i). \tag{1·1}$$

In the theory which takes into account the effect of radiation reaction (Dirac 1939, 1943), the field $\mathbf{A}(\mathbf{x})$ to be used in these Hamiltonian equations is taken to be given by

$$\mathbf{A}(\mathbf{x}) = \tfrac{1}{2}[\mathbf{A}_W(\mathbf{x}+\boldsymbol{\lambda}) + \mathbf{A}_W(\mathbf{x}-\boldsymbol{\lambda})], \tag{1·2}$$

where $\mathbf{A}_W$ is the Wentzel field and $\boldsymbol{\lambda}$ is a small time-like vector which is ultimately made to tend to zero. The $\mathbf{z}_i$ have to satisfy the conditions

$$(\mathbf{z}_i - \mathbf{z}_j \pm \boldsymbol{\lambda})^2 < 0 \quad \text{for} \quad i \neq j. \tag{1·3}$$

The field can be represented by its Fourier-components

$$\mathbf{A}(\mathbf{x}) = \int \mathbf{A}_{\mathbf{k}} e^{i(\mathbf{k},\mathbf{x})} d\mathbf{k}, \quad (d\mathbf{k} = dk_0 dk_1 dk_2 dk_3), \tag{1·4}$$

where
$$\mathbf{A}_{\mathbf{k}} = \frac{1}{(2\pi)^4} \int \mathbf{A}(\mathbf{x})\, e^{-i(\mathbf{k},\mathbf{x})} d\mathbf{x}, \tag{1·5}$$

the reality of $\mathbf{A}(\mathbf{x})$ demanding that $\mathbf{A}_{\mathbf{k}} = \bar{\mathbf{A}}_{-\mathbf{k}}$. The Poisson bracket relations satisfied by the field are taken to be

$$[A_{\mathbf{k}\mu}, A_{\mathbf{k}'\nu}] = -\frac{ig_{\mu\nu}}{4\pi^2} \cos(\mathbf{k},\boldsymbol{\lambda})\, \Delta(\mathbf{k})\, \delta(\mathbf{k}+\mathbf{k}'), \tag{1·6}$$

where
$$\Delta(\mathbf{k}) = \frac{1}{|k|}[\delta(|k| - k_0) - \delta(|k| + k_0)], \quad |k| = (k_1^2 + k_2^2 + k_3^2)^{\frac{1}{2}},$$

$$\delta(\mathbf{k}+\mathbf{k}') = \delta(k_0 + k_0')\, \delta(k_1 + k_1')\, \delta(k_2 + k_2')\, \delta(k_3 + k_3').$$

The Wentzel field satisfies the d'Alembertian equation, which gives

$$\mathbf{k}^2 = 0, \tag{1·7}$$

and a divergence equation, which leads to the (supplementary) condition

$$(\mathbf{A}_{\mathbf{k}}, \mathbf{k}) - \frac{1}{4\pi^2} \cos(\mathbf{k},\boldsymbol{\lambda})\, \Delta(\mathbf{k})\, \Sigma_j e_j e^{-i(\mathbf{k},\mathbf{z}_j)} = 0. \tag{1·8}$$

The usual splitting up of $\mathbf{A}(\mathbf{x})$ into longitudinal and transverse waves is carried out in a particular Lorentz frame of reference. Let the (time-like) vector $\mathbf{a}$ specify the direction of the time-axis of this particular frame in a general Lorentz frame. In terms of Fourier-components, the splitting up is given by

$$A_{\mathbf{k}\mu} = L_{\mathbf{k}\mu} + T_{\mathbf{k}\mu} \equiv L_{\mathbf{k}}^{\mathrm{I}} a_\mu + L_{\mathbf{k}}^{\mathrm{II}} k_\mu + T_{\mathbf{k}\mu}, \tag{1·9}$$

where the transverse components $T_{\mathbf{k}\mu}$ satisfy the relations

$$(\mathbf{T}_{\mathbf{k}}, \mathbf{k}) = 0, \quad (\mathbf{T}_{\mathbf{k}}, \mathbf{a}) = 0. \tag{1·10}$$

It can be seen at once that (1·9) and (1·10) give the familiar splitting up equations if one passes to a frame where $a_\mu = (\mathbf{a}^2)^{\frac{1}{2}} \delta_{\mu 0}$, $\delta_{\mu\nu}$ being the Kronecker symbol.

The equations (1·9) and (1·10) give a general form of the splitting up of the field if $\mathbf{a}$ is regarded as a vector that is arbitrary, apart from satisfying the condition that $(\mathbf{a},\mathbf{k})$ must not vanish identically. This condition must be imposed in order that the equations (1·9) and (1·10) can be solved for $L_{\mathbf{k}}^{\mathrm{I}}$, $L_{\mathbf{k}}^{\mathrm{II}}$ and $T_{\mathbf{k}\mu}$. Thus a different direction $\mathbf{a}$ may be selected for each Fourier-component of the field, say $\mathbf{a}_{\mathbf{k}}$, provided that

$$(\mathbf{a}_{\mathbf{k}},\mathbf{k}) \neq 0. \tag{1·11}$$

The transverse waves so defined are gauge-invariant (as should be the case if they are measurable quantities), since the first equation of (1·10) gives in $\mathbf{x}$-space $\partial T_{\mu}(\mathbf{x})/\partial x_{\mu} = 0$. It is seen from (1·9) and (1·10) that the longitudinal and transverse parts of $\mathbf{A}_{\mathbf{k}}$ remain unchanged if $\mathbf{a}_{\mathbf{k}}$ is replaced in (1·9) by $S'_{\mathbf{k}}\mathbf{a}_{\mathbf{k}}+S''_{\mathbf{k}}\mathbf{k}$ with $S'_{\mathbf{k}} \neq 0$. Introducing the null-vector $\mathbf{l}_{\mathbf{k}}$ defined by

$$\mathbf{l}_{\mathbf{k}} = \frac{\mathbf{a}_{\mathbf{k}}}{(\mathbf{a}_{\mathbf{k}},\mathbf{k})} - \frac{1}{2}\frac{\mathbf{a}_{\mathbf{k}}^{2}\mathbf{k}}{(\mathbf{a}_{\mathbf{k}},\mathbf{k})^{2}}, \tag{1·12}$$

with the properties

$$(\mathbf{l}_{\mathbf{k}},\mathbf{l}_{\mathbf{k}}) = 0, \quad (\mathbf{l}_{\mathbf{k}},\mathbf{k}) = 1, \tag{1·13}$$

one can put the generalized splitting up into the form

$$A_{\mathbf{k}\mu} = L_{\mathbf{k}\mu}+T_{\mathbf{k}\mu} \equiv L_{\mathbf{k}}^{\mathrm{I}} l_{\mathbf{k}\mu} + L_{\mathbf{k}}^{\mathrm{II}} k_{\mu} + T_{\mathbf{k}\mu}, \tag{1·14}$$

where the transverse components satisfy the relations

$$(\mathbf{T}_{\mathbf{k}},\mathbf{l}_{\mathbf{k}}) = 0, \quad (\mathbf{T}_{\mathbf{k}},\mathbf{k}) = 0. \tag{1·15}$$

With the help of (1·13) and (1·15), the equation (1·14) gives the new $L_{\mathbf{k}}^{\mathrm{I}}$ and $L_{\mathbf{k}}^{\mathrm{II}}$ as

$$L_{\mathbf{k}}^{\mathrm{I}} = (\mathbf{A}_{\mathbf{k}},\mathbf{k}), \quad L_{\mathbf{k}}^{\mathrm{II}} = (\mathbf{A}_{\mathbf{k}},\mathbf{l}_{\mathbf{k}}). \tag{1·16}$$

It is not necessary at this stage to restrict $\mathbf{l}_{\mathbf{k}}$ to be real. However, $\mathbf{l}_{\mathbf{k}}$ must satisfy a suitable condition for $L_{\mu}(\mathbf{x})$ (and therefore also $T_{\mu}(\mathbf{x})$) to be real. This condition is $L_{\mathbf{k}\mu} = \bar{L}_{-\mathbf{k}\mu}$, which gives

$$(\mathbf{A}_{\mathbf{k}},\mathbf{k})\,(l_{\mathbf{k}\mu}+\bar{l}_{-\mathbf{k}\mu}) + (\mathbf{A}_{\mathbf{k}},\mathbf{l}_{\mathbf{k}}+\bar{\mathbf{l}}_{-\mathbf{k}})\,k_{\mu} = 0.$$

On multiplication by $l_{\mathbf{k}}^{\mu}$ and by $\bar{l}_{-\mathbf{k}}^{\mu}$, this leads to

$$(\mathbf{l}_{\mathbf{k}},\bar{\mathbf{l}}_{-\mathbf{k}}) = 0, \quad (\mathbf{A}_{\mathbf{k}},\mathbf{l}_{\mathbf{k}}+\bar{\mathbf{l}}_{-\mathbf{k}}) = 0.$$

Therefore it is necessary to impose on permissible functions $\mathbf{l}_{\mathbf{k}}$ the condition that

$$\mathbf{l}_{\mathbf{k}} = -\bar{\mathbf{l}}_{-\mathbf{k}}. \tag{1·17}$$

The Poisson bracket relations of the variables introduced above will now be considered. If $\mathbf{l}_{\mathbf{k}}$ is a function of $\mathbf{k}$ only, it will have vanishing Poisson brackets with all the dynamical variables. More generally, one could take $\mathbf{l}_{\mathbf{k}}$ to be a function either of the co-ordinates $\mathbf{z}_{i}$ or of the momenta $\mathbf{p}_{i}$ of the particles. In either case

$$[l_{\mathbf{k}\mu}, l_{\mathbf{k}'\nu}] = 0, \quad [A_{\mathbf{k}\mu}, l_{\mathbf{k}'\nu}] = 0,$$

$$[L_{\mathbf{k}}^{\mathrm{I}}, l_{\mathbf{k}'\nu}] - 0. \quad [L_{\mathbf{k}}^{\mathrm{II}}, l_{\mathbf{k}'\nu}] = 0 \quad [T_{\mathbf{k}\mu}, l_{\mathbf{k}'\nu}] = 0.$$

From (1·6) and (1·16) one gets, by a straightforward calculation,

$$\left.\begin{aligned}
[L^{\mathrm{I}}_{\mathbf{k}}, L^{\mathrm{I}}_{\mathbf{k}'}] &= 0,\\
[L^{\mathrm{I}}_{\mathbf{k}}, L^{\mathrm{II}}_{\mathbf{k}'}] &= \frac{i}{4\pi^2}\cos(\mathbf{k},\boldsymbol{\lambda})\,\Delta(\mathbf{k})\,\delta(\mathbf{k}+\mathbf{k}'),\\
[L^{\mathrm{II}}_{\mathbf{k}}, L^{\mathrm{II}}_{\mathbf{k}'}] &= -\frac{i}{4\pi^2}(\mathbf{l}_{\mathbf{k}},\mathbf{l}_{\mathbf{k}'})\cos(\mathbf{k},\boldsymbol{\lambda})\,\Delta(\mathbf{k})\,\delta(\mathbf{k}+\mathbf{k}'),\\
[L^{\mathrm{I}}_{\mathbf{k}}, T_{\mathbf{k}'\mu}] &= 0,\\
[L^{\mathrm{II}}_{\mathbf{k}}, T_{\mathbf{k}'\mu}] &= -\frac{i}{4\pi^2}[l_{\mathbf{k}\mu}+l_{\mathbf{k}'\mu}-(\mathbf{l}_{\mathbf{k}},\mathbf{l}_{\mathbf{k}'})\,k_{\mu}]\cos(\mathbf{k},\boldsymbol{\lambda})\,\Delta(\mathbf{k})\,\delta(\mathbf{k}+\mathbf{k}').
\end{aligned}\right\} \tag{1·18}$$

The Poisson brackets of the longitudinal field with the transverse one should vanish (i.e., they should commute in the quantum theory), if they are variables corresponding to different degrees of freedom. The last two equations of (1·18) show that the condition for this is that

$$\mathbf{l}_{\mathbf{k}} = -\mathbf{l}_{-\mathbf{k}}. \tag{1·19}$$

Together with (1·17), (1·19) shows that $\mathbf{l}_{\mathbf{k}}$ must be taken to be a real function of $\mathbf{k}$. This condition and the condition (1·19) will always be assumed to be satisfied in the work that follows. (In terms of $\mathbf{a}_{\mathbf{k}}$ the condition (1·19) is satisfied in particular if

$$a_{\mathbf{k}} = \alpha\mathbf{a}_{-\mathbf{k}}, \tag{1·20}$$

where $\alpha \neq 0$.) With (1·19), one has

$$[L^{\mathrm{II}}_{\mathbf{k}}, L^{\mathrm{II}}_{\mathbf{k}'}] = 0, \quad [L^{\mathrm{II}}_{\mathbf{k}}, T_{\mathbf{k}'\mu}] = 0. \tag{1·21}$$

It will now be assumed that the $\mathbf{l}_{\mathbf{k}}$ depend on $\mathbf{k}$ and on the co-ordinates $\mathbf{z}_i$ ($i = 1, 2, \ldots$). Then $L^{\mathrm{I}}_{\mathbf{k}}$, $L^{\mathrm{II}}_{\mathbf{k}}$ and $\mathbf{T}_{\mathbf{k}}$ depend on the $\mathbf{z}_i$. The following Poisson brackets hold

$$[l_{\mathbf{k}\mu}, z_{i\nu}] = 0, \quad [l_{\mathbf{k}\mu}, p_{i\nu}] = -\frac{\partial l_{\mathbf{k}\mu}}{\partial z^{\nu}_{i}}. \tag{1·22}$$

2. ELIMINATION OF THE LONGITUDINAL WAVES

If all the dynamical variables of the system are transformed to new ones by means of a contact transformation, the new Hamiltonian equations describing the system will be just the old ones expressed in terms of the new variables. It is possible to eliminate the longitudinal waves from the Hamiltonian equations by such a transformation, the supplementary condition (1·8) being at the same time automatically satisfied and so dropping out from the system of field equations. The method of elimination follows that given by Dirac (1943) for ordinary longitudinal waves, adapted to the resolution of the field into Fourier-components.

The required contact transformation can be built up from infinitesimal contact transformations, in which each dynamical variable ξ is transformed according to

$$\frac{d\xi^{(\tau)}}{d\tau} = [\xi^{(\tau)}, G^{(\tau)}] \quad (0 \leqslant \tau \leqslant 1), \tag{2·1}$$

from its initial value $\xi^{(0)} = \xi$ to its final value $\xi^{(1)} = \xi'$ say, with

$$G^{(\tau)} = i\Sigma_j e_j \int (\mathbf{A}_{\mathbf{k}}^{(\tau)}, \mathbf{l}_{\mathbf{k}}^{(\tau)}) \exp [i(\mathbf{k}, \mathbf{z}_j^{(\tau)})]\, d\mathbf{k}. \tag{2·2}$$

From the Poisson brackets of the last section one finds that $\mathbf{l}_{\mathbf{k}}^{(\tau)} = \mathbf{l}_{\mathbf{k}}$ and $\mathbf{z}_i^{(\tau)} = \mathbf{z}_i$. Since $G^{(\tau)}$ is expressible in terms of $L_{\mathbf{k}}^{\mathrm{II}(\tau)}$ (by (1·16)), which has vanishing Poisson brackets with the transverse waves, $\mathbf{T}_{\mathbf{k}}^{(\tau)}$ is also independent of τ. Taking $\xi^{(\tau)} = A_{\mathbf{k}\mu}^{(\tau)}$ in (2·1), one has

$$\begin{aligned} \frac{dA_{\mathbf{k}\mu}^{(\tau)}}{d\tau} &= i\Sigma_j e_j \int [A_{\mathbf{k}\mu}, A_{\mathbf{k}'\rho}]^{(\tau)}\, l_{\mathbf{k}'}^{\rho}\, e^{i(\mathbf{k}', \mathbf{z}_j)}\, d\mathbf{k}' \\ &= -\frac{1}{8\pi^2} \Delta(\mathbf{k})\, l_{\mathbf{k}\mu} \Sigma_j e_j \{e^{-i(\mathbf{k}, \mathbf{z}_j+\lambda)} + e^{-i(\mathbf{k}, \mathbf{z}_j-\lambda)}\}, \end{aligned} \tag{2·3}$$

where (1·6) and (1·19) have been used. By integration with respect to τ this gives

$$A_{\mathbf{k}\mu}^{(\tau)} = A_{\mathbf{k}\mu} - \frac{\tau}{8\pi^2} \Delta(\mathbf{k})\, l_{\mathbf{k}\mu} \Sigma_j e_j \{e^{-i(\mathbf{k}, \mathbf{z}_j+\lambda)} + e^{-i(\mathbf{k}, \mathbf{z}_j-\lambda)}\}. \tag{2·4}$$

The transformation of the $\mathbf{p}_i$ is given by

$$\frac{dp_{i\mu}^{(\tau)}}{d\tau} = -e_i \int (\mathbf{A}_{\mathbf{k}}^{(\tau)}, \mathbf{l}_{\mathbf{k}})\, k_\mu e^{i(\mathbf{k}, \mathbf{z}_i)}\, d\mathbf{k} + i\Sigma_j e_j \int \left(\mathbf{A}_{\mathbf{k}}^{(\tau)}, \frac{\partial \mathbf{l}_{\mathbf{k}}}{\partial z_i^\mu}\right) e^{i(\mathbf{k}, \mathbf{z}_j)}\, d\mathbf{k}. \tag{2·5}$$

The properties (1·13) of $\mathbf{l}_{\mathbf{k}}$ show that $\left(\mathbf{l}_{\mathbf{k}}, \frac{\partial \mathbf{l}_{\mathbf{k}}}{\partial z_i^\mu}\right) = \left(\mathbf{k}, \frac{\partial \mathbf{l}_{\mathbf{k}}}{\partial z_i^\mu}\right) = 0$. Thus, according to (2·4), $\mathbf{A}_{\mathbf{k}}^{(\tau)}$ can be replaced by $\mathbf{A}_{\mathbf{k}}$ in the two integrals of (2·5). Also, from the splitting up equations (1·14) and (1·16), one has

$$(\mathbf{A}_{\mathbf{k}}, \mathbf{l}_{\mathbf{k}})\, k_\mu = A_{\mathbf{k}\mu} - (\mathbf{A}_{\mathbf{k}}, \mathbf{k})\, l_{\mathbf{k}\mu} - T_{\mathbf{k}\mu}$$

and
$$\left(\mathbf{A}_{\mathbf{k}}, \frac{\partial \mathbf{l}_{\mathbf{k}}}{\partial z_i^\mu}\right) = \left(T_{\mathbf{k}}, \frac{\partial \mathbf{l}_{\mathbf{k}}}{\partial z_i^\mu}\right).$$

With the help of these equations, (2·5) becomes

$$\frac{dp_{i\mu}^{(\tau)}}{d\tau} = -e_i A_\mu(\mathbf{z}_i) + e_i T_\mu(\mathbf{z}_i) + e_i \int (\mathbf{A}_{\mathbf{k}}, \mathbf{k})\, l_{\mathbf{k}\mu} e^{i(\mathbf{k}, \mathbf{z}_i)}\, d\mathbf{k} + i\Sigma_j e_j \int \left(\mathbf{T}_{\mathbf{k}}, \frac{\partial \mathbf{l}_{\mathbf{k}}}{\partial z_i^\mu}\right) e^{i(\mathbf{k}, \mathbf{z}_j)}\, d\mathbf{k}. \tag{2·6}$$

After integrating (2·6) over τ from 0 to 1, and expressing $(\mathbf{A}_{\mathbf{k}}, \mathbf{k})$ with the help of the supplementary condition (1·8), one gets

$$\begin{aligned} p_{i\mu} - e_i A_\mu(\mathbf{z}_i) = p'_{i\mu} - e_i T_\mu(\mathbf{z}_i) &- i\Sigma_j e_j \int \left(\mathbf{T}_{\mathbf{k}}, \frac{\partial \mathbf{l}_{\mathbf{k}}}{\partial z_i^\mu}\right) e^{i(\mathbf{k}, \mathbf{z}_j)}\, d\mathbf{k} \\ &- \frac{e_i}{8\pi^2} \Sigma_j e_j \int \Delta(\mathbf{k})\, l_{\mathbf{k}\mu} \{e^{i(\mathbf{k}, \mathbf{z}_i-\mathbf{z}_j+\lambda)} + e^{i(\mathbf{k}, \mathbf{z}_i-\mathbf{z}_j-\lambda)}\}\, d\mathbf{k}. \end{aligned} \tag{2·7}$$

The left-hand side of (2·7) is the expression (1·1), in function of which the Hamiltonian equation of the ith particle is written (both in the classical and in the quantum theory). The right-hand side of (2·7) is expressed in terms of the new dynamical variables introduced by the contact transformation; the longitudinal field variables have been eliminated and replaced by the two terms containing integrals. The first integral contains in general transverse field variables referring to *all* the particles (not only to the ith one), but it vanishes if $\partial l_{\mathbf{k}\nu}/\partial z_i^\mu = 0$. The second integral depends on the co-ordinates of the particles and the choice of $\mathbf{l}_\mathbf{k}$, but not on any field variables. It is interesting to examine when the term corresponding to $j = i$ ('self-energy') in the summation over j in this integral vanishes. This term is of the form

$$\int \Delta(\mathbf{k})\, \mathbf{l}_\mathbf{k} \cos(\mathbf{k}, \boldsymbol{\lambda})\, d\mathbf{k} = \iiint \frac{1}{|k|} \mathbf{l}_\mathbf{k}^+ \cos b\, |k|\, dk_1\, dk_2\, dk_3 - \iiint \frac{1}{|k|} \mathbf{l}_\mathbf{k}^- \cos c\, |k|\, dk_1\, dk_2\, dk_3, \tag{2·8}$$

where $\mathbf{l}_\mathbf{k}^+$, $\mathbf{l}_\mathbf{k}^-$ is $\mathbf{l}_\mathbf{k}$ with k_0 replaced by $\pm |k|$ respectively,

$$b = \lambda_0 - |\lambda| \cos\beta, \quad c = \lambda_0 + |\lambda| \cos\beta$$

and β is the angle between the 3-vectors λ_s and k_s. Since $\boldsymbol{\lambda}$ is a time-like vector, b and c are different from zero. Now transform the right-hand side of (2·8) to polar co-ordinates $r = |k|$, θ, ϕ, and suppose $\mathbf{l}_\mathbf{k}^+$ and $\mathbf{l}_\mathbf{k}^-$ expansible in power series of r with powers $\geqslant -1$ in the form $\mathbf{l}_\mathbf{k}^+ = \sum_0^\infty \mathbf{b}_n r^{2n-1}$, $\mathbf{l}_\mathbf{k}^- = \sum_0^\infty \mathbf{c}_n r^{2n-1}$ with $\mathbf{b}_n$ and $\mathbf{c}_n$ independent of r. These series can only contain odd powers of r due to the condition (1·19). The integrals over r in (2·8) become

$$\int_0^\infty \sum_0^\infty \mathbf{b}_n r^{2n} \cos br\, dr, \quad \int_0^\infty \sum_0^\infty \mathbf{c}_n r^{2n} \cos cr\, dr. \tag{2·9}$$

Each of the two integrals of (2·9) may be evaluated by term by term integration, whenever this procedure, applied to the integral in question and to the corresponding integral obtained by replacing cos by sin, leads to a series that is summable (in the Cesàro sense) (Hardy 1908). Since

$$\int_0^\infty r^{2n} \cos br\, dr = 0, \quad \int_0^\infty r^{2n} \sin br\, dr = (-1)^n (2n)!/b^{2n+1},$$

the 'self-energy' vanishes if $\sum_0^\infty (-1)^n (2n)!\, \mathbf{b}_n/b^{2n+1}$ and $\sum_0^\infty (-1)^n (2n)!\, \mathbf{c}_n/c^{2n+1}$ are summable. In particular, the 'self-energy' vanishes if $\mathbf{a}_\mathbf{k}$ depends only on the direction of $\mathbf{k}$, so that $\mathbf{a}_{m\mathbf{k}} = \alpha \mathbf{a}_\mathbf{k}$ for any m, where m and α are non-vanishing scalars, since the power series above reduced to their respective first terms in this case.

If the $\mathbf{l}_\mathbf{k}$ are such that the corresponding $\mathbf{a}_\mathbf{k}$ may be taken independent of $\mathbf{k}$, the last term of (2·7) gives, after passing to the limit $\boldsymbol{\lambda} \to 0$,

$$-\frac{e_i}{2} \sum_{j \neq i} e_j \frac{[2(\mathbf{a}, \mathbf{z}_i - \mathbf{z}_j)^2 - \mathbf{a}^2(\mathbf{z}_i - \mathbf{z}_j)^2]\, a_\mu - (\mathbf{a}, \mathbf{z}_i - \mathbf{z}_j)\, \mathbf{a}^2 (z_{i\mu} - z_{j\mu})}{[(\mathbf{a}, \mathbf{z}_i - \mathbf{z}_j)^2 - \mathbf{a}^2(\mathbf{z}_i - \mathbf{z}_j)^2]^{\frac{3}{2}}}. \tag{2·10}$$

If the $\mathbf{a}$ are also independent of all the $\mathbf{z}_i$, the first integral on the right-hand side of (2·7) vanishes, and (2·10) agrees with the result given by Dirac (1943) in a reference frame where $a_\mu = (\mathbf{a}^2)^{\frac{1}{2}} \delta_{\mu 0}$. In the single-time Hamiltonian theory, (2·10) then leads to the electrostatic Coulomb potential of the particles.

One could also consider the case where the $\mathbf{l}_\mathbf{k}$ are functions of $\mathbf{k}$ and of the $\mathbf{p}_i$. The contact transformation (2·1) with (2·2) then makes the new $\mathbf{z}_i$ depend on the $\mathbf{A}_\mathbf{k}$, thus leading to difficulties. This case is therefore abandoned.

Up to this point, each step taken in the classical theory above can be carried over into the quantum theory. The right-hand side of the supplementary condition (1·8) must be applied onto the wave-function. One arrives at the same result (2·7) with both sides of that equation applied onto the wave-function.

3. CLASSICAL DEFLEXION OF AN ELECTRON BY A NUCLEUS

As an application of the general theory above, consider the problem of the interaction of two charged particles, say an electron and a nucleus, in the classical theory.

The usual method, where the field is split up with reference to a selected time-axis, taken to be the direction of the initial velocity 4-vector of the nucleus, leads to the Coulomb interaction. If the interaction with the transverse field and the nuclear recoil are neglected, the following (relativistic) formula for the angle of deflexion ϕ of an electron, which is supposed coming from infinity, is obtained (Darwin 1913; Schonland 1922)

$$tg[\tfrac{1}{2}(\pi+\phi)\sqrt{(1-(b_0/b)^2)}] = \frac{b}{b_0}\frac{v_1}{\sqrt{(1+v_1^2)}}\sqrt{(1-(b_0/b)^2)}. \tag{3·1}$$

Here ϕ is taken to be positive when the deflexion is towards the nucleus, $b_0 = Ee/mv_1$, E = charge of the nucleus, $-e$ = charge of the electron, m = rest-mass of the electron, v_1 = spatial part of the initial 4-velocity of the electron, and b is the impact parameter. For small angles of deflexion one can put $\sqrt{(1-(b_0/b)^2)} \approx 1$, and (3·1) gives

$$\phi \approx \frac{2Ee}{mb}\frac{\sqrt{(1+v_1^2)}}{v_1^2} \approx \frac{2Ee}{mbv_1} \quad \text{if} \quad v_1 \gg 1. \tag{3·2}$$

A different method of splitting up the field will now be introduced. The initial velocity 4-vectors of the nucleus and of the electron determine a set of parallel two-dimensional planes in space-time, which is taken as reference for defining the $\mathbf{a}_\mathbf{k}$. These may be defined by the following geometrical method.

Only such functions $\mathbf{a}_\mathbf{k}$ will be considered whose direction depends only on the direction of $\mathbf{k}$, so that $\mathbf{a}_{m\mathbf{k}} = \alpha\mathbf{a}_\mathbf{k}$ for any m, where m and α are non-vanishing scalars. The direction of $\mathbf{a}_\mathbf{k}$ (which determines the splitting up) is specified by its point at infinity, $A(K)$ say, where K is the point at infinity of the direction of $\mathbf{k}$. All the points at infinity of space-time form a 3-dimensional projective space S^3_∞. The $g_{\mu\nu}$-metric in space-time defines a null-quadric in S^3_∞. The selected set of parallel planes has a line r in S^3_∞, that intersects the null-quadric. The problem of defining the direction of $\mathbf{a}_\mathbf{k}$ is equivalent to the problem of defining in S^3_∞ a point-function $A(K)$, where K is a variable point on the null-quadric, by means of projective operations, the line r being used as reference. Since the $\mathbf{a}_\mathbf{k}$ have to satisfy (1·11), the point $A(K)$ must not lie on the polar plane of K. The condition (1·20) is automatically satisfied.

The function $A(K)$ may be defined as follows. Consider the plane $[rK]$ through the line r and the point K (figure 1). The polar line of r, $\bar{r}$ say, does not lie on this plane and it therefore intersects this plane in a point, O say. Take $A(K) = O$ for all the points K of the null-quadric which lie on the plane $[rK]$.

To express $A(K)$ analytically, introduce homogeneous co-ordinates X_μ, X_4 ($\mu = 0, 1, 2, 3$) into space-time by putting $x_\mu = X_\mu/X_4$. The points of S^3_∞ have $X_4 = 0$, and the null-quadric is given by $X_\mu X^\mu \equiv (\mathbf{X},\mathbf{X}) = 0$. Let P, Q be a pair of polar points on r, such that P lies outside and Q inside the null-quadric. The co-ordinates of P and Q satisfy the equation $(\mathbf{P},\mathbf{Q}) = 0$. The equation of $\bar{r}$ is obtained by looking upon $\bar{r}$ as the intersection of the polar planes of P and Q, namely $(\mathbf{P},\mathbf{X}) = 0$ and $(\mathbf{Q},\mathbf{X}) = 0$. Any point of the plane $[rK] \equiv [PQK]$ may be represented parametrically by $X_\mu = \gamma_1 P_\mu + \gamma_2 Q_\mu + \gamma_3 K_\mu$. The point O in which $\bar{r}$ intersects the plane $[PQK]$ can therefore be obtained from the solution of the equations

$$O_\mu = \gamma_1 P_\mu + \gamma_2 Q_\mu + \gamma_3 K_\mu, \quad (\mathbf{P},\mathbf{O}) = 0, \quad (\mathbf{Q},\mathbf{O}) = 0.$$

Hence
$$O_\mu = (\mathbf{P},\mathbf{K})(\mathbf{Q},\mathbf{Q}) P_\mu + (\mathbf{Q},\mathbf{K})(\mathbf{P},\mathbf{P}) Q_\mu - (\mathbf{P},\mathbf{P})(\mathbf{Q},\mathbf{Q}) K_\mu.$$

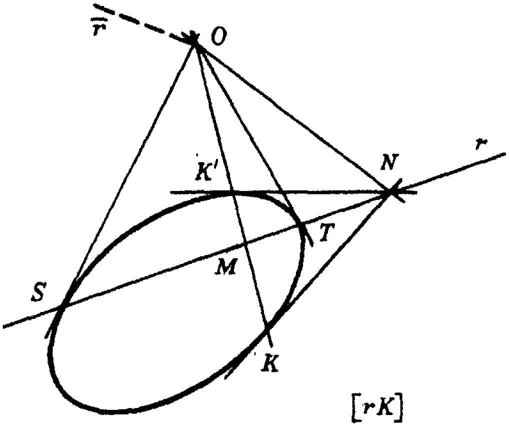


FIGURE 1

On returning to non-homogeneous co-ordinates in space-time, and dropping the part parallel to **k**, which makes no change in the splitting up of the field, this gives

$$\mathbf{a}_\mathbf{k} = (\mathbf{p},\mathbf{k})\mathbf{p} - (\mathbf{q},\mathbf{k})\mathbf{q}, \tag{3·3}$$

where
$$\mathbf{p}^2 = -1, \quad \mathbf{q}^2 = 1, \quad (\mathbf{p},\mathbf{q}) = 0. \tag{3·4}$$

The choice of $\mathbf{a}_\mathbf{k}$ above is the only simple one. For instance, the intersection of OK with r, $[OK, r] = M$ say, taken as $A(K)$ leads to the same result (3·3), and so does the choice as $A(K)$ of any point on the line OK. Further, if S, T are the two points where r intersects the null-quadric, one might think of taking the harmonic conjugate N of M with respect to S, T as $A(K)$, but this is not permissible since N can be shown to lie on the polar plane of K. Other choices for $\mathbf{a}_\mathbf{k}$ are more complicated and artificial, and so are not of interest.

Take a Lorentz frame of reference in which the nucleus is initially at rest at the origin, with the z_1-axis parallel to the spatial part of the initial 4-velocity of the electron, and the z_3-axis orthogonal in three-space to the plane of motion. The points at infinity of the z_0-axis and the z_1-axis are two polar points on r, which can be taken as the points Q and P introduced above, so that $q_\mu = \delta_{\mu 0}$ and $p_\mu = \delta_{\mu 1}$ in (3·3). Then (1·12) gives

$$l_{\mathbf{k}\mu} = (k_0 \delta_{\mu 0} + k_1 \delta_{\mu 1} - \tfrac{1}{2} k_\mu)/(k_0^2 - k_1^2). \tag{3·5}$$

This defines the new way of splitting up the field. The new longitudinal field is replaced in the Hamiltonian equations by the last term on the right-hand side of (2·7), as the $\mathbf{l}_k$ of (3·5) are independent of the $\mathbf{z}_i$. The 'self-energy' term $j = i$ vanishes as discussed earlier. One may now put $\boldsymbol{\lambda} = 0$ in (2·7), and pass to the single-time Hamiltonian theory by putting the time variables of the two particles equal, $z_{i0} = z_0$, and adding the Hamiltonian equations of the two particles in the usual way to form the total Hamiltonian function. The three spatial components of the last term in (2·7) vanish. Thus the last term of (2·7) contributes the following 'interaction potential' $\mathscr{V}$ to the total Hamiltonian

$$\begin{aligned}\mathscr{V} &= \frac{1}{8\pi^2}\sum_i e_i \sum_{j \neq i} e_j \int \Delta(\mathbf{k})\,\frac{k_0}{k_0^2 - k_1^2}\,e^{-i\Sigma_t k_t(z_{it}-z_{jt})}\,d\mathbf{k} \quad (t = 1, 2)\\ &= -\frac{Ee}{\pi}\,\delta(z_1 - Z_1)\int_0^{2\pi} d\phi \int_0^{\infty} \frac{1}{r}\,e^{-i|z_2 - Z_2| r\cos\phi}\,dr, \qquad (3{\cdot}6)\end{aligned}$$

where the co-ordinates of the nucleus are denoted by capital letters, and those of the electron by small letters. The integral over r in (3·6) diverges near the point $r = 0$ independently of $z_2 - Z_2$, since

$$\int_0^{\infty} \frac{1}{r}\,e^{-izr\cos\phi}\,dr = \int_0^{\rho} \frac{1}{r}\,dr + \int_{\rho}^{\infty} \frac{1}{r}\,e^{-izr\cos\phi}\,dr,$$

where $\rho > 0$ is small. Now

$$\frac{\partial}{\partial z}\int_{\rho}^{\infty} \frac{1}{r}\,e^{-\epsilon r - izr\cos\phi}\,dr = -i\int_{\rho}^{\infty} \cos\phi\, e^{-\epsilon r - izr\cos\phi}\,dr \to -\frac{1}{z},$$

where $e^{-\epsilon r}$ is a Cesàro convergence factor. Thus the constant infinity can be separated out in (3·6), and one gets

$$\mathscr{V} = 2Ee\,\delta(z_1 - Z_1)\,[\log|z_2 - Z_2| + \text{const.}] \qquad (3{\cdot}7)$$

as the expression replacing the Coulomb potential. It will be found that the infinite constant has no physical effect on the motion of the particles.

On the assumption that the interaction with the new transverse field may be neglected, the equations of motion of the two particles are as follows, where S and s are the proper-times of the nucleus and of the electron respectively,

$$M\frac{d^2Z_t}{dS^2} = -\frac{dz_0}{dS}\frac{\partial\mathscr{V}}{\partial Z_t}, \quad m\frac{d^2z_t}{ds^2} = -\frac{dz_0}{ds}\frac{\partial\mathscr{V}}{\partial z_t} \quad (t = 1, 2), \qquad (3{\cdot}8)$$

$$M\frac{d^2z_0}{dS^2} = -\Sigma_t\frac{dZ_t}{dS}\frac{\partial\mathscr{V}}{\partial Z_t}, \quad m\frac{d^2z_0}{ds^2} = -\Sigma_t\frac{dz_t}{ds}\frac{\partial\mathscr{V}}{\partial z_t}, \qquad (3{\cdot}9)$$

and the trivial equations

$$\left(\frac{dz_0}{dS}\right)^2 - \left(\frac{dZ_1}{dS}\right)^2 - \left(\frac{dZ_2}{dS}\right)^2 = 1, \quad \left(\frac{dz_0}{ds}\right)^2 - \left(\frac{dz_1}{ds}\right)^2 - \left(\frac{dz_2}{ds}\right)^2 = 1. \qquad (3{\cdot}10)$$

The system (3·8), (3·9) and (3·10) may be regarded as giving Z_t and z_t ($t = 1, 2$) as functions of z_0. Let z_0^* be the value of z_0 for which $z_1 - Z_1 = 0$. Since the potential $\mathscr{V}$ contains a δ-function, which vanishes everywhere except at z_0^*, the motion of the two particles is straightline and uniform before and after z_0^*. In a relativistic theory, where the velocity of the particles is always finite, they will change their direction of motion by a finite angle at the instant z_0^* due to the δ-potential. It may be pointed out that with such a δ-potential it would be physically wrong to neglect the nuclear recoil. The two sets of equations (3·8) give the following momentum equations

$$M\frac{dZ_1}{dS}+m\frac{dz_1}{ds}=mv_1, \tag{3·11}$$

where v_1 is the spatial part of the initial 4-velocity of the electron, and

$$M\frac{dZ_2}{dS}+m\frac{dz_2}{ds}=0. \tag{3·12}$$

The two equations (3·9) give the conservation of energy equation

$$M\frac{dz_0}{dS}+m\frac{dz_0}{ds}=\mathscr{V}+M+m\sqrt{(1+v_1^2)}. \tag{3·13}$$

The second equation of (3·8) for $t=2$ is

$$m\frac{d^2z_2}{ds^2}=-\frac{2Ee}{z_2-Z_2}\delta(z_1-Z_1)\frac{dz_0}{ds}. \tag{3·14}$$

Try a solution of the form

$$\frac{dz_2}{ds}=v_2'\epsilon(z_1-Z_1), \tag{3·15}$$

where v_2' is a constant and

$$\epsilon(z_1-Z_1)=\begin{cases}0 & \text{for}\quad -\infty<z_1-Z_1<0,\\ 1 & \text{for}\quad 0<z_1-Z_1<\infty.\end{cases}$$

The differentiation of (3·15) gives

$$\begin{aligned}\frac{d^2z_2}{ds^2}&=v_2'\delta(z_1-Z_1)\left\{\frac{dz_1}{ds}-\frac{dZ_1}{dS}\frac{dz_0/ds}{dz_0/dS}\right\}\\ &=v_2'\delta(z_1-Z_1)\left\{\frac{dz_1}{ds}\left(1+\frac{m}{M}\frac{dz_0/ds}{dz_0/dS}\right)-\frac{m}{M}\frac{dz_0/ds}{dz_0/dS}v_1\right\},\end{aligned} \tag{3·16}$$

where (3·11) has been used. In a small neighbourhood of z_0^*, the equations (3·10), (3·11) and (3·12) give

$$\frac{dz_0/ds}{dz_0/dS}=\left\{\frac{1+(dz_1/ds)^2+(dz_2/ds)^2}{1+\frac{m^2}{M^2}(v_1-dz_1/ds)^2+\frac{m^2}{M^2}(dz_2/ds)^2}\right\}^{\frac{1}{2}}\sim\frac{dz_1/ds}{\frac{m}{M}\,|\,v_1-dz_1/ds\,|}\sim\frac{M}{m}.$$

Omitting in (3·16) the term which remains finite in a small neighbourhood of z_0^*, one has

$$\frac{d^2z_2}{ds^2}\sim 2v_2'\delta(z_1-Z_1)\frac{dz_0}{ds}, \tag{3·17}$$

so that, from (3·14),

$$v_2'=-Ee/mb, \tag{3·18}$$

where b is the impact parameter. The z_1-component v_1' of the final 4-velocity of the electron is obtained from the conservation of energy

$$M\sqrt{(1+V_1'^2+V_2'^2)}+m\sqrt{(1+v_1'^2+v_2'^2)}=M+m\sqrt{(1+v_1^2)},$$

where V_1' and V_2' are the z_1- and z_2-components of the final 4-velocity of the nucleus, namely,

$$V_1' = \frac{m}{M}(v_1 - v_1'), \quad V_2' = -\frac{m}{M}v_2'.$$

Thus v_1' satisfies the equation

$$\begin{aligned}[][m^2v_1^2 - (M + m\sqrt{(1+v_1^2)})^2]\,v_1'^2 + 2m(m + M\sqrt{(1+v_1^2)})\,v_1v_1' \\ + (m + M\sqrt{(1+v_1^2)})^2 - (1+v_2'^2)\,(M + m\sqrt{(1+v_1^2)})^2 = 0. \quad (3\cdot19)\end{aligned}$$

The angle of deflexion ϕ of the electron, taken to be positive when the deflexion is towards the nucleus, is given by

$$\sin\phi = \frac{Ee}{mb}\frac{1}{\sqrt{(v_1'^2 + v_2'^2)}}, \qquad (3\cdot20)$$

where v_1' satisfies (3·19) and v_2' is given by (3·18). The deflexion (3·20) can never exceed a right angle, as can be seen directly from the nature of the potential $\mathscr{V}$.

If $M \gg m$, the equation (3·19) gives as first approximation $v_1' = \sqrt{(v_1^2 - v_2'^2)}$, and (3·20) becomes

$$\sin\phi = Ee/mbv_1 \approx \phi$$

if ϕ is small. This is half the deflexion (3·2) derived from the Coulomb potential for ϕ small and $v_1 \gg 1$. Thus, for small velocities v_1, where the deflexion (3·1) agrees with experiment, the method proposed above of splitting up the field and neglecting the new transverse waves does not lead to agreement with experiment, which means that the effect of the new transverse waves is not negligible for small velocities of the electron. However, for a very high energy electron, the usual method of working only with the Coulomb interaction force involves neglecting radiation damping in a certain way, and in this case the new method proposed above may be a better one.

In conclusion, I wish to express my deepest gratitude to Professor P. A. M. Dirac, F.R.S. for his supervision and guidance during the course of the above work. I am also indebted to the British Council for the award of a scholarship.

REFERENCES

Darwin, C. G. 1913 *Phil. Mag.* **25**, 201.
Dirac, P. A. M. 1939 *Ann. Inst. Poincaré*, **9**, 13.
Dirac, P. A. M. 1943 *Comm. Dubl. Inst. Adv. Studies*, **A**, no. 1.
Fermi, E. 1930 *R.C. Accad. Lincei*, **12**, 431.
Fermi, E. 1932 *Rev. Mod. Phys.* **4**, 87.
Hardy, G. H. 1908 *Trans. Camb. Phil. Soc.* **21**, 1.
Schonland, B. F. J. 1922 *Proc. Roy. Soc.* **A**, **101**, 299.

The laminar boundary-layer equations of bodies of revolution. Motion of a sphere

BY D. MEKSYN, D.SC.

(*Communicated by G. Temple, F.R.S.—Received* 17 *September* 1947)

1. INTRODUCTION AND SUMMARY

In a previous paper (Meksyn 1948) the author gave a method of calculating the velocity distribution in a laminar boundary layer on cylindrical bodies. The aim of the present paper is to extend the method to the case of laminar boundary layers on bodies of revolution.

The problem has been treated, so far, only in a few papers (Goldstein 1938, §§ 51, 52, 61).

Millikan (1932) has derived boundary-layer equations and Kármán's momentum equation for bodies of revolution, and, assuming a parabolic distribution for the velocity u in the laminar part, and $u \sim y^{1/7}$ in the turbulent part (where y is the distance from the surface), he has applied the momentum equation to airship-line bodies.

Fediaersky (1934) derived independently the momentum equation for bodies of revolution and applied it to airship-like bodies by assuming $u \sim u^n$, where n is arbitrary.

Tomotika (1935) applied the momentum equation to the evaluation of various boundary-layer quantities for bodies of revolution, in particular for a sphere.

Fage (1936) used Tomotika's results to evaluate the surface friction at the forward part of a sphere.

In the present paper the general boundary-layer equations are derived for axially symmetrical motions of bodies of revolution; the equations are partially solved for a sphere, and the results compared with Fage's (1936) measurements.

The main results obtained are as follows:

(1) The problem leads to a partial differential equation (2·20) which for the front part of the body reduces to the well-known equation

$$f'''+ff'' = \lambda(1-f'^2),$$

where λ is a parameter.

(2) The evaluated surface friction agrees for the front part of the sphere with Fage's computations and measurements.

In the case $R = 251{,}300$ the agreement between the evaluated and observed values is satisfactory up to 85° from the forward stagnation point.

(3) The pressure consists of two parts; one is equal to the inviscid pressure, the other one depends on the thickness of the boundary layer.

A. GENERAL EQUATIONS OF MOTION OF THE BOUNDARY LAYER

2. Equations of motion

It was shown in the previous paper (Meksyn 1948) that the motion in the boundary layer can be described by a stream function of the orthogonal co-ordinates (α, β), where α is the inviscid velocity potential and β the corresponding stream function; namely, by a function $\psi(\alpha, \beta)$ which has the following properties.

The second derivative of ψ with respect to β is of the form

$$\psi''(\alpha, \beta) \sim A(\alpha, \beta)\, e^{\xi(\alpha, \beta)},$$

where $A(\alpha, \beta)$ and $e^{\xi(\alpha, \beta)}$ are slowly and rapidly varying functions respectively; differentiation with respect to β increases the order of magnitude of ψ'' by $R^{\frac{1}{2}}$, whereas differentiation with respect to α leaves the order of magnitude unchanged; the order of ψ'' is equal to $R^{\frac{1}{2}}$, where R is an appropriate Reynolds number.

These results will now be applied to motion with axial symmetry.

The equations of motion in general co-ordinates are as follows (Goldstein 1938, p. 114).

Let α, β and ϖ be orthogonal curvilinear co-ordinates, where α and β are in the meridian plane, whereas ϖ is proportional to the azimuthal angle, and let the element of length ds be

$$ds^2 = h_1^2 d\alpha^2 + h_2^2 d\beta^2 + h_3^2 d\varpi^2. \tag{2·1}$$

In what follows the motion is independent of ϖ.

The equation of continuity is

$$\frac{\partial}{\partial \alpha}(h_2 h_3 u) + \frac{\partial}{\partial \beta}(h_3 h_1 v) = 0, \tag{2·2}$$

where u and v are the velocities along α and β; there is accordingly a stream function ψ such that

$$h_3 u = \frac{1}{h_2}\frac{\partial \psi}{\partial \beta}, \quad h_3 v = -\frac{1}{h_1}\frac{\partial \psi}{\partial \alpha}, \tag{2·3}$$

the vorticity is

$$\left.\begin{aligned} \zeta &= -\frac{1}{h_3} D^2\psi, \\ \text{where} \quad D^2 &= \frac{h_3}{h_1 h_2}\left[\frac{\partial}{\partial \alpha}\left(\frac{h_2}{h_3 h_1}\frac{\partial}{\partial \alpha}\right) + \frac{\partial}{\partial \beta}\left(\frac{h_1}{h_2 h_3}\frac{\partial}{\partial \beta}\right)\right], \end{aligned}\right\} \tag{2·4}$$

and the equation of motion is

$$-\frac{1}{h_1 h_2 h_3}\frac{\partial(\psi, D^2\psi)}{\partial(\alpha, \beta)} + \frac{2D^2\psi}{h_1 h_2 h_3^2}\frac{\partial(\psi, h_3)}{\partial(\alpha, \beta)} \doteq \nu D^4\psi. \tag{2·5}$$

Expanding, and retaining only terms of the two highest orders, we find that

$$\begin{aligned} D^2\psi &= \frac{1}{h_2^2}\frac{\partial^2\psi}{\partial \beta^2} + \frac{\partial \psi}{\partial \beta}\frac{h_3}{h_1 h_2}\frac{\partial}{\partial \beta}\left(\frac{h_1}{h_2 h_3}\right), \\ D^4\psi &= \frac{1}{h_2^4}\left\{\frac{\partial^4\psi}{\partial \beta^4} + 2\frac{\partial^3\psi}{\partial \beta^3}\frac{\partial}{\partial \beta}\log(h_1 h_3^{-1} h_2^{-3})\right\}, \\ \frac{\partial}{\partial \beta}D^2\psi &= \frac{1}{h_2^2}\left\{\frac{\partial^3\psi}{\partial \beta^3} + \frac{\partial^2\psi}{\partial \beta^2}\frac{\partial \log(h_1 h_2^{-3} h_3^{-1})}{\partial \beta}\right\}, \\ \frac{\partial}{\partial \alpha}D^2\psi &= \frac{1}{h_2^2}\left\{\frac{\partial^3\psi}{\partial \alpha\, \partial \beta^2} + \frac{\partial^2\psi}{\partial \beta^2}\frac{\partial \log h_2^{-2}}{\partial \alpha} + \frac{\partial^2\psi}{\partial \beta\, \partial \alpha}\frac{\partial}{\partial \beta}\log(h_1 h_2^{-1} h_3^{-1})\right. \\ &\qquad \left. + \frac{\partial \psi}{\partial \beta}\left[\frac{\partial^2 \log(h_1 h_2^{-1} h_3^{-1})}{\partial \alpha\, \partial \beta} + \frac{\partial \log h_2^{-2}}{\partial \alpha}\frac{\partial \log(h_1 h_2^{-1} h_3^{-1})}{\partial \beta}\right]\right\}. \end{aligned} \tag{2·6}$$

Substituting (2·6) in (2·5), we obtain the following result:

$$\begin{aligned}
-\frac{1}{h_1h_2h_3}&\left\{\underset{R^{1/2}}{\frac{\partial\psi}{\partial\alpha}}\frac{1}{h_2^2}\left[\underset{R^{2}}{\frac{\partial^3\psi}{\partial\beta^3}}+\underset{R^{3/2}}{\frac{\partial^2\psi}{\partial\beta^2}}\frac{\partial\log(h_1h_2^{-3}h_3^{-1})}{\partial\beta}\right]-\underset{R}{\frac{\partial\psi}{\partial\beta}}\frac{1}{h_2^2}\left[\underset{R^{3/2}}{\frac{\partial^3\psi}{\partial\alpha\,\partial\beta^2}}+\underset{R^{3/2}}{\frac{\partial^2\psi}{\partial\beta^2}}\frac{\partial\log h_2^{-2}}{\partial\alpha}\right.\right.\\
&\left.\left.+\underset{R}{\frac{\partial^2\psi}{\partial\beta\,\partial\alpha}}\frac{\partial}{\partial\beta}\log(h_1h_2^{-1}h_3^{-1})+\underset{R}{\frac{\partial\psi}{\partial\beta}}\left(\frac{\partial^2\log(h_1h_2^{-1}h_3^{-1})}{\partial\alpha\,\partial\beta}+\frac{\partial\log h_2^{-2}}{\partial\alpha}\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right)\right]\right\}\\
&+\frac{2}{h_1h_2^3h_3^2}\left[\underset{R^{3/2}}{\frac{\partial^2\psi}{\partial\beta^2}}+\underset{R}{\frac{\partial\psi}{\partial\beta}}\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right]\left[\underset{R^{1/2}}{\frac{\partial\psi}{\partial\alpha}}\frac{\partial h_3}{\partial\beta}-\underset{R}{\frac{\partial\psi}{\partial\beta}}\frac{\partial h_3}{\partial\alpha}\right]\\
&\qquad\qquad=\frac{\nu}{h_2^4}\left\{\underset{R^{5/2}}{\frac{\partial^4\psi}{\partial\beta^4}}+2\underset{R^{2}}{\frac{\partial^3\psi}{\partial\beta^3}}\frac{\partial}{\partial\beta}\log(h_1h_3^{-1}h_2^{-3})\right\},
\end{aligned}\tag{2·7}$$

where underneath are indicated the orders of magnitude of the terms, the order of the h's and their derivatives being unity.

Retaining the terms of orders $R^{5/2}$ and R^2, and rearranging we find that

$$\begin{aligned}
\nu\frac{h_1h_3}{h_2}\frac{\partial^4\psi}{\partial\beta^4}&+\frac{\partial\psi}{\partial\alpha}\frac{\partial^3\psi}{\partial\beta^3}-\frac{\partial\psi}{\partial\beta}\frac{\partial^3\psi}{\partial\alpha\,\partial\beta^2}-\frac{\partial\psi}{\partial\beta}\frac{\partial^2\psi}{\partial\beta^2}\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}\\
&=-2\nu\frac{\partial^3\psi}{\partial\beta^3}\frac{h_1h_3}{h_2}\frac{\partial}{\partial\beta}\log(h_1h_3^{-1}h_2^{-3})-\frac{\partial\psi}{\partial\alpha}\frac{\partial^2\psi}{\partial\beta^2}\frac{\partial}{\partial\beta}\log(h_1h_2^{-3}h_3^{-3})\\
&\quad+\left(\frac{\partial\psi}{\partial\beta}\right)^2\left[\frac{\partial^2\log(h_1h_2^{-1}h_3^{-1})}{\partial\alpha\,\partial\beta}+\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right]\\
&\quad+\frac{\partial\psi}{\partial\beta}\frac{\partial^2\psi}{\partial\beta\,\partial\alpha}\frac{\partial}{\partial\beta}\log(h_1h_2^{-1}h_3^{-1}).
\end{aligned}\tag{2·8}$$

On expansion in powers of β we obtain the equation

$$\left.\begin{aligned}
\frac{h_1h_3}{h_2}&=\left(\frac{h_1h_3}{h_2}\right)_0+\left(\frac{\partial}{\partial\beta}\frac{h_1h_3}{h_2}\right)_0\beta,\\
\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}&=\left(\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}\right)_0+\left(\frac{\partial^2\log(h_2^{-2}h_3^{-2})}{\partial\alpha\,\partial\beta}\right)_0\beta,
\end{aligned}\right\}\tag{2·9}$$

where the subscript denotes that the expression is evaluated on the boundary $\beta=0$.

Combining (2·8) and (2·9), and integrating with respect to β, from infinity to β, we see that

$$\begin{aligned}
\nu\left(\frac{h_1h_3}{h_2}\right)_0\frac{\partial^3\psi}{\partial\beta^3}&+\frac{\partial\psi}{\partial\alpha}\frac{\partial^2\psi}{\partial\beta^2}-\frac{\partial\psi}{\partial\beta}\frac{\partial^2\psi}{\partial\alpha\,\partial\beta}-\frac{1}{2}\left(\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}\right)_0\left[\left(\frac{\partial\psi}{\partial\beta}\right)^2-q^2\right]\\
&=-2\nu\frac{\partial^2\psi}{\partial\beta^2}\left(\frac{h_1h_3}{h_2}\right)_0\left(\frac{\partial\log(h_1h_3^{-1}h_2^{-3})}{\partial\beta}\right)_0+\int_\infty^\beta\left\{-\nu\frac{\partial^4\psi}{\partial\beta^4}\left(\frac{h_1h_3}{h_2}\right)_0\left(\frac{\partial\log(h_1h_3h_2^{-1})}{\partial\beta}\right)_0\beta\right.\\
&\quad+\frac{\partial\psi}{\partial\beta}\frac{\partial^2\psi}{\partial\beta^2}\left(\frac{\partial^2\log(h_2^{-2}h_3^{-2})}{\partial\alpha\,\partial\beta}\right)_0\beta-\frac{\partial\psi}{\partial\alpha}\frac{\partial^2\psi}{\partial\beta^2}\left(\frac{\partial\log(h_1h_2^{-3}h_3^{-3})}{\partial\beta}\right)_0\\
&\quad\left.+\frac{\partial\psi}{\partial\beta}\frac{\partial^2\psi}{\partial\alpha\,\partial\beta}\left(\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right)_0\right\}d\beta+\int^\beta\left(\frac{\partial\psi}{\partial\beta}\right)^2\left[\left(\frac{\partial^2\log(h_1h_2^{-1}h_3^{-1})}{\partial\alpha\,\partial\beta}\right)_0\right.\\
&\quad\left.+\left(\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}\right)_0\left(\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right)_0\right]d\beta.
\end{aligned}\tag{2·10}$$

The upper limit of the integral of $(\partial\psi/\partial\beta)^2$ is left indeterminate; it should be taken at the edge of the boundary layer, while q has the value of $\partial\psi/\partial\beta$ as β tends to infinity.

To find q, let the flow be in the direction of x positive, the velocity at infinity being u_0; in what follows α and β are the velocity potential and the stream function of the corresponding inviscid motion referred to speed unity at infinity.

Consider the co-ordinates x, y in the meridian plane, u and v being the corresponding velocities respectively; now

$$yu = \frac{\partial\psi}{\partial y} = \frac{\partial\psi}{\partial\alpha}\frac{\partial\alpha}{\partial y} + \frac{\partial\psi}{\partial\beta}\frac{\partial\beta}{\partial y}. \tag{2·11}$$

Since

$$\frac{\partial\alpha}{\partial y} = 0, \quad y = \frac{\partial\beta}{\partial y} \quad \text{at} \quad y = \infty, \tag{2·12}$$

it follows that

$$u_0 = q = \frac{\partial\psi}{\partial\beta} \quad \text{at} \quad y = \infty. \tag{2·13}$$

We now introduce a new co-ordinate γ, such that

$$l + \alpha = \gamma, \tag{2·14}$$

where l is a length chosen so that

$$\gamma = 0 \quad \text{at the forward stagnation point.} \tag{2·15}$$

We also write

$$\sigma = \frac{1}{2}\left(\frac{2u_0}{\nu\gamma}\right)^{\frac{1}{2}}\beta, \quad \psi = (2\nu u_0\gamma)^{\frac{1}{2}} f(\sigma, \gamma). \tag{2·16}$$

From (2·16) we obtain the relations

$$\left.\begin{aligned}
&\frac{\partial\psi}{\partial\beta} = u_0\frac{\partial f}{\partial\sigma}, \quad \frac{\partial\psi}{\partial\alpha} = \frac{1}{2}\left(\frac{2\nu u_0}{\gamma}\right)^{\frac{1}{2}}\left(f - \sigma\frac{\partial f}{\partial\sigma} + 2\gamma\frac{\partial f}{\partial\gamma}\right), \quad \frac{\partial^2\psi}{\partial\beta^2} = \frac{u_0}{2}\left(\frac{2u_0}{\nu\gamma}\right)^{\frac{1}{2}}\frac{\partial^2 f}{\partial\sigma^2}, \\
&\frac{\partial^2\psi}{\partial\alpha\,\partial\beta} = \frac{u_0}{2\gamma}\left(-\sigma\frac{\partial^2 f}{\partial\sigma^2} + 2\gamma\frac{\partial^2 f}{\partial\sigma\partial\gamma}\right), \quad \frac{\partial^3\psi}{\partial\beta^3} = \frac{u_0^2}{2\nu\gamma}\frac{\partial^3 f}{\partial\sigma^3}, \quad \frac{\partial^4\psi}{\partial\beta^4} = \frac{u_0^2}{4\nu\gamma}\left(\frac{2u_0}{\nu\gamma}\right)^{\frac{1}{2}}\frac{\partial^4 f}{\partial\sigma^4}.
\end{aligned}\right\} \tag{2·17}$$

Substituting (2·17) in (2·10)

$$\begin{aligned}
&\left(\frac{h_1h_3}{h_2}\right)_0\frac{\partial^3 f}{\partial\sigma^3} + f\frac{\partial^2 f}{\partial\sigma^2} + 2\gamma\left[\frac{\partial f}{\partial\gamma}\frac{\partial^2 f}{\partial\sigma^2} - \frac{\partial f}{\partial\sigma}\frac{\partial^2 f}{\partial\gamma\,\partial\sigma}\right]\gamma\left(\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}\right)_0\left[1 - \left(\frac{\partial f}{\partial\sigma}\right)^2\right] \\
&\quad = 2\left(\frac{\nu}{2u_0\gamma}\right)^{\frac{1}{2}}\left\{-2\gamma\left(\frac{\partial\log(h_1h_3^{-1}h_2^{-3})}{\partial\beta}\right)_0\left(\frac{h_1h_3}{h_2}\right)_0\frac{\partial^2 f}{\partial\sigma^2}\right. \\
&\qquad + \int_\infty^\sigma\left[-\left(\frac{h_1h_3}{h_2}\right)_0\gamma\left(\frac{\partial\log(h_1h_3/h_2)}{\partial\beta}\right)_0\sigma\frac{\partial^4 f}{\partial\sigma^4} + 2\gamma^2\left(\frac{\partial^2\log(h_2^{-2}h_3^{-2})}{\partial\alpha\,\partial\beta}\right)_0\sigma\frac{\partial f}{\partial\sigma}\frac{\partial^2 f}{\partial\sigma^2}\right. \\
&\qquad - \gamma\left(\frac{\partial\log(h_1h_2^{-3}h_3^{-3})}{\partial\beta}\right)_0\frac{\partial^2 f}{\partial\sigma^2}\left(f - \sigma\frac{\partial f}{\partial\sigma} + 2\gamma\frac{\partial f}{\partial\gamma}\right) + \gamma\left(\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right)_0 \\
&\qquad \left.\times\frac{\partial f}{\partial\sigma}\left(-\sigma\frac{\partial^2 f}{\partial\sigma^2} + 2\gamma\frac{\partial^2 f}{\partial\sigma\,\partial\gamma}\right)\right]d\sigma + \int^\sigma 2\gamma^2\left[\left(\frac{\partial^2\log(h_1h_2^{-1}h_3^{-1})}{\partial\alpha\,\partial\beta}\right)_0\right. \\
&\qquad \left.\left. + \left(\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}\right)_0\left(\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right)_0\right]\left(\frac{\partial f}{\partial\sigma}\right)^2 d\sigma\right\}.
\end{aligned} \tag{2·18}$$

Write

$$\kappa\sigma_1 = \sigma, \quad f(\sigma, \gamma) = \kappa f_1(\sigma_1, \gamma), \quad \kappa = \left(\frac{h_1h_3}{h_2}\right)^{\frac{1}{2}}. \tag{2·19}$$

Then (2·18) becomes, after few transformations, and dropping the subscripts

$$\frac{\partial^3 f}{\partial\sigma^3}+f\frac{\partial^2 f}{\partial\sigma^2}+2\gamma\left[\frac{\partial f}{\partial\gamma}\frac{\partial^2 f}{\partial\sigma^2}-\frac{\partial f}{\partial\sigma}\frac{\partial^2 f}{\partial\gamma\,\partial\sigma}\right]+\gamma\left(\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}\right)_0\left[1-\left(\frac{\partial f}{\partial\sigma}\right)^2\right]$$

$$=2\left(\frac{\nu}{2u_0\gamma}\right)^{\frac{1}{2}}\left(\frac{h_1h_3}{h_2}\right)^{\frac{1}{2}}\left\{-\gamma\left(\frac{\partial\log(h_1h_3h_2^{-1})}{\partial\beta}\right)_0\sigma\frac{\partial^3 f}{\partial\sigma^3}+\gamma\left(\frac{\partial\log(h_3^3h_2^5h_1^{-1})}{\partial\beta}\right)_0\frac{\partial^2 f}{\partial\sigma^2}\right.$$

$$+\int_\infty^\sigma\left[2\gamma^2\left(\frac{\partial^2\log(h_2^{-2}h_3^{-2})}{\partial\alpha\,\partial\beta}\right)_0-\gamma\left(\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right)_0\sigma\frac{\partial f}{\partial\sigma}\frac{\partial^2 f}{\partial\sigma^2}\right.$$

$$\left.+\gamma\left(\frac{\partial\log(h_1h_2^{-3}h_3^{-3})}{\partial\beta}\right)_0\frac{\partial^2 f}{\partial\sigma^2}\left(f-\sigma\frac{\partial f}{\partial\sigma}+2\gamma\frac{\partial f}{\partial\gamma}\right)-\gamma\left(\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right)_0 2\frac{\partial f}{\partial\sigma}\gamma\frac{\partial^2 f}{\partial\gamma\,\partial\sigma}\right]d\sigma$$

$$\left.+\int^\sigma 2\gamma^2\left[\left(\frac{\partial^2\log(h_1h_2^{-1}h_3^{-1})}{\partial\alpha\,\partial\beta}\right)_0+\left(\frac{\partial\log(h_2^{-2}h_3^{-2})}{\partial\alpha}\right)_0\left(\frac{\partial\log(h_1h_2^{-1}h_3^{-1})}{\partial\beta}\right)_0\right]\left(\frac{\partial f}{\partial\sigma}\right)^2 d\sigma\right\}, \tag{2·20}$$

where
$$\sigma=\frac{1}{2}\left(\frac{2u_0}{\nu\gamma}\right)^{\frac{1}{2}}\left(\frac{h_2}{h_1h_3}\right)_0^{\frac{1}{2}}\beta \quad\text{and}\quad \psi=(2\nu u_0\gamma)^{\frac{1}{2}}\left(\frac{h_1h_3}{h_2}\right)^{\frac{1}{2}}f(\sigma,\gamma). \tag{2·21}$$

Since the velocity vanishes on the boundary, the function $f(\sigma,\gamma)$ has to satisfy the following conditions (2·13)

$$f=0=\frac{\partial f}{\partial\sigma}\quad\text{at}\quad\sigma=0,\qquad\frac{\partial f}{\partial\sigma}=1\quad\text{at}\quad\sigma=\infty. \tag{2·22}$$

In the case of a sphere the coefficient of the term

$$\int^\sigma\left(\frac{\partial f}{\partial\sigma}\right)^2 d\sigma$$

vanishes.

For the front part of the body, the equation (2·20) can be considerably simplified, as it will appear below, and it becomes

$$\frac{\partial^3 f}{\partial\sigma^3}+f\frac{\partial^2 f}{\partial\sigma^2}=-\gamma\left(\frac{\partial\log h_2^{-3}h_3^{-2}}{\partial\alpha}\right)_0\left[1-\left(\frac{\partial f}{\partial\sigma}\right)^2\right], \tag{2·23}$$

which is identical with the equation in the case of a cylindrical body (Meksyn 1948).

3. FRICTION

The surface friction is found from the equation

$$F=-\mu\zeta=\frac{\mu D^2\psi}{h_3}\equiv\frac{\mu}{h_3h_2^2}\frac{\partial^2\psi}{\partial\beta^2}, \tag{3·1}$$

where only the term of the highest order of magnitude is retained.

Making use of (2·17), transforming to the new co-ordinate σ_1 (2·19) and dropping the subscripts, we obtain

$$\frac{F}{\frac{1}{2}\rho u_0^2}=4\left(\frac{\nu}{2u_0\gamma}\right)^{\frac{1}{2}}\frac{1}{h_3h_2^2}\left(\frac{h_2}{h_1h_3}\right)^{\frac{1}{2}}\frac{\partial^2 f}{\partial\sigma^2}, \tag{3·2}$$

where f is found from (2·20).

4. PRESSURE

Pressure is found from the tensor equation of motion (Goldstein 1938, p. 115).

$$\frac{\partial \mathbf{v}}{\partial t}-\mathbf{v}\times\boldsymbol{\omega} = -\operatorname{grad}\left(\frac{p}{\rho}+\tfrac{1}{2}\mathbf{v}^2\right)-\nu\operatorname{curl}\boldsymbol{\omega}, \tag{4·1}$$

where $\mathbf{v}$ and $\boldsymbol{\omega}$ are the vectors of velocity and vorticity respectively.

Now
$$\operatorname{curl}\boldsymbol{\omega} = \frac{1}{h_2 h_3}\frac{\partial}{\partial \beta}(h_3\zeta), \tag{4·2}$$

and from (2·4) we find, after few transformations, that

$$h_3\zeta = -\frac{1}{h_2^3 h_3}\left\{\frac{\partial^3\psi}{\partial\beta^3}+\left(\frac{\partial}{\partial\beta}\log\frac{h_1}{h_2^3 h_3}\right)_0\frac{\partial^2\psi}{\partial\beta^2}\right\}, \tag{4·3}$$

where in (4·3) only the two terms of the highest order are retained.

From (4·1) and (4·3), if p is the pressure on the boundary, we obtain the equation

$$\frac{1}{\rho}\frac{1}{h_1}\frac{\partial p}{\partial\alpha} = \frac{\nu}{h_2^3 h_3}\left\{\frac{\partial^3\psi}{\partial\beta^3}+\left(\frac{\partial}{\partial\beta}\log\frac{h_1}{h_2^3 h_3}\right)_0\frac{\partial^2\psi}{\partial\beta^2}\right\}. \tag{4·4}$$

Transforming (4·4) to the co-ordinate σ_1, by making use of (2·17) and (2·19), we get

$$\frac{1}{\frac{1}{2}\rho u_0^2}\frac{\partial p}{\partial\alpha} = \frac{1}{\gamma}\frac{1}{h_2^2 h_3^2}f_1'''(0)+2\left(\frac{\nu}{2u_0\gamma}\right)^{\frac{1}{2}}\frac{h_1}{h_2^3 h_3}\left(\frac{h_1 h_3}{h_2}\right)^{-\frac{1}{2}}\left(\frac{\partial\log h_1 h_2^{-3}h_3^{-1}}{\partial\beta}\right)_0 f_1''(0,\gamma) \tag{4·5}$$

where primes indicate differentiation with respect to σ.

Substituting instead of $f_1'''(0,\gamma)$ its value from (2·20), and dropping the subscripts, we get

$$\begin{aligned}\frac{2}{\rho u_0^2}\frac{\partial p}{\partial\alpha} = {}& -\frac{\partial}{\partial\alpha}\frac{1}{h_2^2 h_3^2}+\frac{2}{\gamma h_2^2 h_3^2}\left(\frac{2u_0\gamma}{\nu}\right)^{-\frac{1}{2}}\left(\frac{h_1 h_3}{h_2}\right)^{\frac{1}{2}}\left\{-\gamma\left(\frac{\partial\log h_2^{-2}h_3^{-2}}{\partial\beta}\right)_0\frac{\partial^2 f}{\partial\sigma^2}\right.\\ &+\int_\infty^0\left[\left(2\gamma^2\left(\frac{\partial^2\log h_2^{-2}h_3^{-2}}{\partial\alpha\,\partial\beta}\right)_0-\gamma\left(\frac{\partial\log h_1 h_2^{-1}h_3^{-1}}{\partial\beta}\right)_0\right)\sigma\frac{\partial f}{\partial\sigma}\frac{\partial^2 f}{\partial\sigma^2}\right.\\ &-\gamma\left(\frac{\partial\log h_1 h_2^{-3}h_3^{-3}}{\partial\beta}\right)_0\frac{\partial^2 f}{\partial\sigma^2}\left(f-\sigma\frac{\partial f}{\partial\sigma}+2\gamma\frac{\partial f}{\partial\gamma}\right)\\ &\left.+\gamma\left(\frac{\partial\log h_1 h_2^{-1}h_3^{-1}}{\partial\beta}\right)_0 2\frac{\partial f}{\partial\sigma}\gamma\frac{\partial^2 f}{\partial\sigma\,\partial\gamma}\right]d\sigma+\int^0 2\gamma^2\left[\left(\frac{\partial^2\log h_1 h_2^{-1}h_3^{-1}}{\partial\alpha\,\partial\beta}\right)_0\right.\\ &\left.\left.+\left(\frac{\partial\log h_2^{-2}h_3^{-2}}{\partial\alpha}\right)_0\left(\frac{\partial\log h_1 h_2^{-1}h_3^{-1}}{\partial\beta}\right)_0\right]\left(\frac{\partial f}{\partial\sigma}\right)^2 d\sigma\right\},\end{aligned} \tag{4·6}$$

where in (4·6) all relevant quantities are taken on the boundary $\sigma = 0$, and in the last integral the second limit depends on the thickness of the boundary layer.

The first term on the right-hand side (4·6) can be integrated, whence

$$\frac{2}{\rho u_0}p_1 = \text{const.}-h_2^{-2}h_3^{-2}. \tag{4·7}$$

The remaining terms depend on the thickness of the boundary layer and have to be integrated numerically.

B. MOTION OF A SPHERE

5. PRELIMINARY FORMULAE

Consider the laminar motion round a sphere of radius a; the velocity of the undisturbed flow is in the direction of x positive, and is equal to u_0.

Denoting the usual co-ordinates in the meridian plane by r and θ, we let the velocity potential and the stream function, corresponding to the speed unity at infinity, be α and β respectively; then (Lamb 1932, pp. 123, 129)

$$\alpha = r\cos\theta + \frac{1}{2}\frac{a^3}{r^2}\cos\theta, \quad \beta = \frac{1}{2}\frac{r^2}{a}\sin^2\theta - \frac{1}{2}\frac{a^2}{r}\sin^2\theta, \tag{5·1}$$

where β is divided by a in order to make the dimensions of α and β equal to that of length; α and β are orthogonal co-ordinates.

To find h_1 and h_2, consider an element of length in the meridian plane

$$ds^2 = dr^2 + r^2 d\theta^2 = \left[\left(\frac{\partial r}{\partial\alpha}\right)^2 + r^2\left(\frac{\partial\theta}{\partial\alpha}\right)^2\right]d\alpha^2 + \left[\left(\frac{\partial r}{\partial\beta}\right)^2 + r^2\left(\frac{\partial\theta}{\partial\beta}\right)^2\right]d\beta^2 = h_1^2 d\alpha^2 + h_2^2 d\beta^2, \tag{5·2}$$

where the term proportional to $d\alpha\, d\beta$ vanishes, since α and β are orthogonal.

Differentiating (5·1) and solving with respect to dr and $d\theta$ we find that

$$\left.\begin{aligned}
&\frac{\partial r}{\partial\alpha} = -\frac{(-1+a^3/r^3)\,r/a\sin\theta\cos\theta}{\Delta}, \qquad \frac{\partial r}{\partial\beta} = \frac{(1+a^3/2r^3)\sin\theta}{\Delta},\\
&r\frac{\partial\theta}{\partial\alpha} = -\frac{(1+\frac{1}{2}a^3/r^3)\,r/a\sin^2\theta}{\Delta}, \qquad r\frac{\partial\theta}{\partial\beta} = -\frac{(-1+a^3/r^3)\cos\theta}{\Delta},\\
&\Delta = \left(-1+\frac{a^3}{r^3}\right)^2\frac{r}{a}\cos^2\theta\sin\theta + \left(1+\frac{a^3}{2r^3}\right)^2\frac{r}{a}\sin^3\theta,
\end{aligned}\right\} \tag{5·3}$$

whence from (5·2) and (5·3)

$$\left.\begin{aligned}
h_1^2 &= \frac{(-1+a^3/r^3)^2\,r^2/a^2\sin^2\theta\cos^2\theta + (1+\frac{1}{2}a^3/r^3)^2\,r^2/a^2\sin^4\theta}{\Delta^2},\\
h_2^2 &= \frac{(1+a^3/2r^3)^2\sin^2\theta + (-1+a^3/r^3)^2\cos^2\theta}{\Delta^2}.
\end{aligned}\right\} \tag{5·4}$$

The third co-ordinate ϖ can be taken to be

$$\varpi = a\phi, \tag{5·5}$$

where ϕ is the azimuthal angle; and, since an element of length perpendicular to the meridian plane is equal to $r\sin\theta\, d\phi$, it follows that

$$h_3 = \frac{r}{a}\sin\theta. \tag{5·6}$$

The co-ordinate γ (2·14) is given by

$$\gamma = l + \alpha = \frac{3a}{2} + r\cos\theta + \frac{1}{2}\frac{a^3}{r^2}\cos\theta, \tag{5·7}$$

and at the forward stagnation point

$$\theta = \pi, \quad \gamma = 0,$$

while on the boundary $r = a$

$$\gamma_0 = 3a\cos^2\tfrac{1}{2}\theta. \tag{5·8}$$

We have now to evaluate the differential coefficients of the h's with respect to α and β; the computations are very simple; as an example one term will be considered, and then the final results will be given.

From (5·3) on the boundary $r = a$

$$\frac{\partial r}{\partial \alpha} = 0, \quad \frac{\partial r}{\partial \beta} = \frac{2}{3\sin^2\theta}, \quad a\frac{\partial \theta}{\partial \alpha} = -\frac{2}{3\sin\theta}, \quad a\frac{\partial \theta}{\partial \beta} = 0. \tag{5·9}$$

Consider the term $\partial h_1/\partial\beta$. Since

$$\frac{\partial \log h_1^2}{\partial \beta} = \frac{\partial \log h_1^2}{\partial r}\frac{\partial r}{\partial \beta} + \frac{\partial \log h_1^2}{\partial \theta}\frac{\partial \theta}{\partial \beta} = \frac{\partial \log h_1^2}{\partial r}\frac{\partial r}{\partial \beta}, \tag{5·10}$$

therefore from (5·4) and (5·9)

$$\frac{\partial \log h_1^2}{\partial \beta} = \frac{\partial \log h_3^2}{\partial \beta} = \frac{4}{3a\sin^2\theta}, \quad \text{while} \quad \frac{\partial h_2}{\partial \beta} = 0. \tag{5·11}$$

The evaluation of the differential coefficients with respect to α is even simpler, since we can put in (5·4) $r = a$ before differentiation.

From (5·4), (5·8) and (5·11) we obtain

$$\left.\begin{aligned}\gamma_0\left(\frac{\partial}{\partial \alpha}\log h_2^2 h_3^2\right)_0 = \frac{\cos\theta}{\sin^2\frac{1}{2}\theta}, \quad \gamma_0\left(\frac{\partial \log h_1^2}{\partial \beta}\right)_0 = \gamma_0\left(\frac{\partial \log h_3^2}{\partial \beta}\right)_0 = \frac{1}{\sin^2\frac{1}{2}\theta},\\ \gamma_0\left(\frac{\partial \log h_2^2}{\partial \beta}\right)_0 = 0, \quad \gamma_0^2\left(\frac{\partial^2 \log h_1^2}{\partial \alpha\,\partial \beta}\right)_0 = \frac{\cos\theta}{\sin^4\frac{1}{2}\theta}, \quad \left(\frac{h_1 h_3}{h_2}\right)_0^{\frac{1}{2}} = \sin\theta.\end{aligned}\right\} \tag{5·12}$$

Making use of (5·12), we can easily evaluate the relevant expressions in (2·20).

6. Friction

The equation (2·20) will be considered, but only without the terms depending on the thickness of the boundary layer and on the partial differential coefficient with respect to γ, namely,

$$f''' + ff'' = \gamma_0\left(\frac{\partial}{\partial \alpha}\log h_2^2 h_3^2\right)_0 (1 - f'^2) = \lambda(1 - f'^2),$$

where

$$\lambda = f'''(0) = \frac{\cos\theta}{\sin^2\frac{1}{2}\theta}. \tag{6·1}$$

The surface friction F (3·2) is accordingly found from the equation

$$\frac{1}{\frac{1}{2}\rho u_0^2}F = 6\sqrt{3}\left(\frac{2au_0}{\nu}\right)^{-\frac{1}{2}} f''(0)\sin^2\tfrac{1}{2}\theta\cos\tfrac{1}{2}\theta. \tag{6·2}$$

7. Pressure

Making use of (4·6) the pressure can be evaluated; it is equal to

$$p = p_1 + p_2, \tag{7·1}$$

where (4·7)

$$\frac{1}{\frac{1}{2}\rho u_0^2}p_1 = \text{const.} - \frac{1}{h_2^2 h_3^2} = \tfrac{9}{8}\cos 2\theta - \tfrac{1}{8} \tag{7·2}$$

the constant of integration being found from the condition at the forward stagnation point

$$\theta = \pi, \quad p_1 = \frac{\rho u_0^2}{2},$$

p_1 is equal to the inviscid pressure.

The component p_2, depending on the thickness of the boundary layer, is found from (4·6), (5·12) in the form

$$\frac{1}{\frac{1}{2}\rho u_0^2}\frac{\partial p_2}{\partial \theta} = -6\sqrt{3}\left(\frac{2au_0}{\nu}\right)^{-\frac{1}{2}}\{[f''(0)+A-B-2C]\sin\theta\sin\tfrac{1}{2}\theta+4A\cos\theta\cos\tfrac{1}{2}\theta\}, \quad (7{\cdot}3)$$

where
$$A = \int_0^\infty \sigma f'f''\,d\sigma, \quad B = \int_0^\infty ff''\,d\sigma, \quad C = \int_0^\infty \gamma\frac{\partial f}{\partial \gamma}\frac{\partial^2 f}{\partial \sigma^2}\,d\sigma. \quad (7{\cdot}4)$$

In the above transformations use was made of the relation

$$\frac{\partial}{\partial \alpha} = \frac{\partial \theta}{\partial \alpha}\frac{\partial}{\partial \theta} = -\frac{2}{3a\sin\theta}\frac{\partial}{\partial \theta}. \quad (7{\cdot}5)$$

p_2 can be evaluated by numerical integration, the constant of integration being found from the condition at the forward stagnation point

$$\theta = \pi, \quad p_2 = 0.$$

It should be noticed that p_2 refers to the pressure on the surface of the sphere and f refers to the equation (2·20); the integrals (7·4) are functions of θ.

8. COMPARISON WITH EXPERIMENTAL RESULTS

Pressure and surface friction were measured by Fage (1936) on a sphere for several Reynolds numbers within and beyond the critical range; the corresponding data are given in table 1.

TABLE 1

u_0 (ft./sec.)	$R = 2u_0a/\nu$	
50	157200	separation at 83°
80	251300	—
95	298500	—
135	424500	beyond critical range

In the last three cases (table 1) the boundary layer became turbulent before the separation point.

Separation

The separation point is found from the condition $f''(0) = 0$, which corresponds to the vanishing of surface friction.

Making use in the first approximation of the simplified equation (6·1), the separation point is found from the condition (Hartree 1937)

$$\lambda = \frac{\cos\theta}{\sin^2\frac{1}{2}\theta} = 0{\cdot}199,$$

which corresponds to $\theta = 84°\,48'$, i.e. $95°\,12'$ from the forward stagnation point, exactly as in the case of a circular cylinder.

The terms disregarded in (2·20) depend on the thickness of the boundary layer and also on the differential coefficients of f with respect to the second variable γ; if the former terms only were taken into account the separation point would be slightly advanced.

The observed separation point in the case $u_0 = 50$ ft./sec. was about 83° from the forward stagnation point; in the other cases the boundary layer became turbulent before separation.

In the case $R = 251,300$, if the curve of the observed laminar surface friction is continued, it intersects the axis at about 91° from the forward stagnation point.

Friction

The surface friction F is obtained from (6·2). To find F for a given angle, the quantity $\lambda = f'''(0)$ (6·1) is first evaluated, and the corresponding value of $f''(0)$ can be taken from Hartree's (1937) numerical solution.

The relevant values at $f''(0)$ and $f'''(0)$ are given for the sake of convenience in table 2.

Table 2

$-\lambda$	−0·18	−0·16	−0·14	−0·10	0	0·1	0·2
$f''(0)$	0·128	0·190	0·239	0·319	0·4696	0·587	0·687
$-\lambda$	0·3	0·4	0·5	0·6	0·8	1·0	
$f''(0)$	0·775	0·854	0·928	0·996	1·120	1·233	

The surface friction was measured only within the separation and transition regions; close to the stagnation point the friction was evaluated by Fage (1936) by using Tomotika's method.

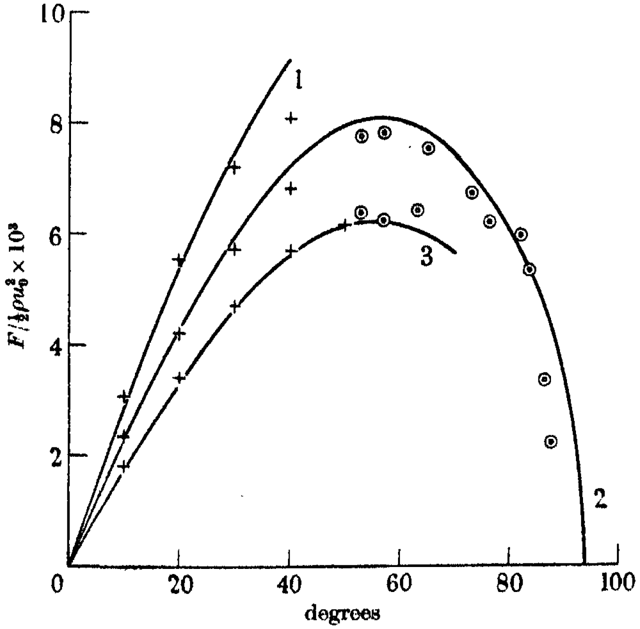


Figure 1. + calc. by Fage; ⊙ obs. by Fage.
u_0 values in ft./sec. are: curve 1, 50; curve 2, 80; curve 3, 135.

In figure 1 are plotted the theoretical values of the surface friction together with the values evaluated and measured by Fage.

As can be seen from figure 1 there is a close agreement between our values and those calculated by Fage.

There is good agreement between the theoretical and observed friction, for the case $u_0 = 80$ ft./sec., up to 85° from stagnation point.

The present results are in general very similar to those in the case of a circular cylinder.

I wish to thank Professor G. Temple, F.R.S., for many helpful suggestions.

REFERENCES

Fage, A. 1936 *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1766.
Fediaevsky, K. 1934 *Trans. Central Aero-Hydrodyn. Inst. Moscow*, no. 179.
Goldstein, S. 1938 *Modern developments in fluid dynamics.* Oxford: Clarendon Press.
Hartree, D. R. 1937 *Proc. Camb. Phil. Soc.* **33**, 223.
Lamb, H. 1932 *Hydrodynamics.* Camb. Univ. Press.
Meksyn, D. 1948 *Proc. Roy. Soc.* A, **192**, 545.
Millikan, C. B. 1932 *Trans. Amer. Soc. Mech. Engrs* (Appl. Mech. Sec.), **54**, 29.
Tomotika, S. 1935 *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1678.

On the scattering of protons by protons

BY W. H. RAMSEY, *Turner and Newall Fellow, University of Manchester*

(*Communicated by P. M. S. Blackett, F.R.S.—Received* 20 *November* 1947
—*Revised* 19 *January* 1948)

The features of the scattering of protons by protons are calculated using the Møller-Rosenfeld version of the meson theory of nuclear forces. A comparison with the experimental material suggests that the neutron-proton interaction is slightly stronger than that between two protons. The predictions of the theory fit the experimental results best when the meson mass is assumed to be about 270 electronic masses. This is significantly larger than that indicated by cosmic-ray measurements, and it is also greater than the values found by using the theory to calculate the features of neutron-proton scattering and the binding energies of the lightest nuclei.

INTRODUCTION

The technique of measuring the absolute cross-section for the scattering of protons by protons has been developed to a high degree of precision, and reliable experimental data are available over a wide range of energies. This material can provide valuable information about the nuclear interaction between two protons, a detailed study of which is not possible by any other means.

In the present paper the interaction between two nucleons will be described in terms of the symmetrical version of the meson theory of nuclear forces which has been developed by Møller & Rosenfeld (1940). The static interaction in this theory contains three adjustable parameters, two of which will be chosen so that the energies of the stable and virtual states of the deuteron are given correctly; the third parameter is the mass of the meson. The empirical data will be used to check the predicted angular distribution of the scattered protons and to determine the mass of the meson. Special attention will be paid to the relationship between the neutron-proton and proton-proton interactions, both as regards magnitude and range of action.

Recently, Wilson (1947) has investigated proton-proton scattering at an energy of 10 MeV, and he has detected a *P*-wave anomaly in addition to the *S*-wave anomaly known hitherto. His results indicate that the nuclear interaction between two protons in a ^{3}P state is weakly repulsive, a feature which is in accord with the version of meson theory adopted in this paper. This experiment, which gives the first information on the nuclear interaction in a ^{3}P state, provides an independent test of a theory of nuclear forces. Wilson's results will be analyzed in detail and compared with the quantitative predictions of the Møller-Rosenfeld theory.

PRELIMINARY

In the mixed symmetrical meson theory of Møller & Rosenfeld (1940) the static part of the interaction between two nucleons is given by

$$\mathbf{V} = (\mathbf{G}/r)\, e^{-\lambda r}, \tag{1}$$

where r is the distance between the nucleons and where the inverse range λ is related to the meson mass m by the equation

$$\lambda = mc/\hbar. \tag{2}$$

The reasons for the adoption of this particular version of the meson theory of nuclear forces have been presented by the author in a previous paper (1947). The eigenvalue of the operator **G** depends only on the symmetry properties of the wave function describing the behaviour of the two nucleons; its value is determined if it is specified whether the state is a singlet or a triplet and whether the azimuthal quantum number l is even or odd.

The Møller-Rosenfeld theory predicts a strong repulsion between two nucleons in a ^{1}P state, and a weak repulsion in a ^{3}P state. Proton-proton scattering cannot give any information about the former interaction as the ^{1}P configuration is forbidden by the Pauli principle. The validity of the theory in this connexion has, however, already been established by the results of experiments on the scattering of fast neutrons by protons (cf. Ramsey 1947). The absence of the strong ^{1}P interaction in the proton-proton case opens up the possibility of studying the interaction in the ^{3}P state, which is weakly repulsive in the Møller-Rosenfeld scheme. This feature of the theory is confirmed, at least qualitatively, by recent experiments by Wilson (1947) on the scattering of 10 MeV protons by protons.

The static interaction of the theory adopted contains three adjustable parameters, the mass of the meson and two charge constants. For a fixed meson mass the charge constants will be chosen so that the theory leads to the correct energy levels for the stable and virtual states of the deuteron. Ramsey (1947) has already fixed the charges in this manner but, for the purposes of the present paper, it has been necessary to extend the calculations to higher values of the meson mass. It is convenient to replace the interaction constant G by the dimensionless parameter

$$b = -MG/m\hbar c, \tag{3}$$

where M is the mass of a proton. The results of the calculations are given in table 1 (m_e denotes the mass of an electron). The uncertainties indicated in the values of b allow for the errors in the experimental data on the states of the deuteron and for errors inherent in the method of calculation; it is the former which gives the major contribution. The restrictions placed on b are stringent and its value is known to 0·04 % for the ^{3}S state and to 0·2 % for the ^{1}S state. The latter uncertainty may seem surprisingly low in view of the fact that the empirical result on which it is based, namely, the cross-section for the scattering of thermal neutrons by protons, is known only to 5 %. This feature is closely connected with the fact that, were the cross-section infinite so that the virtual ^{1}S state has zero energy, the value of b would be increased by only 10 % to 1·68 (cf. Hulthén 1944).

TABLE 1. VALUES OF CONSTANT b

meson mass	^{3}S state	^{1}S state
$225m_e$	$2{\cdot}543 \pm 0{\cdot}001$	$1{\cdot}530 \pm 0{\cdot}004$
$275m_e$	$2{\cdot}389 \pm 0{\cdot}001$	$1{\cdot}555 \pm 0{\cdot}004$
$300m_e$	$2{\cdot}331 \pm 0{\cdot}001$	$1{\cdot}565 \pm 0{\cdot}004$

In addition to (1), there is a non-static interaction whose influence on the energy levels of the deuteron is appreciable (cf. Møller & Rosenfeld 1940). This difficulty can be removed, as Rosenfeld (1945) has pointed out, by a modification of the method of separating the total interaction into static and non-static parts. The non-static interaction then contains only non-central terms and so has no influence in the ^{1}S configuration, either directly or through the determination of the constant b*; it is this state which controls proton-proton scattering at low energies. On the other hand, the value of b for the ^{3}S state is liable to slight modification on account of the non-static interaction, which also gives rise to the quadrupole moment of the deuteron: this is not, however, important in the present paper. Likewise the addition of a non-central term to the static interaction would not affect the subsequent discussion.

METHODS OF CALCULATION

General accounts of the wave mechanical treatment of the scattering of protons by protons have been given by Breit, Condon & Present (1936) and by Rosenfeld (1947). The problem can be reduced to the evaluation of a set of phase shifts η_l by a standard procedure.

The simplest method to determine a phase shift is to use the Born approximation, but this is reliable only if the energy is very high or if the quantum number l is large.

---

* Theoretically the non-static interaction has an influence on the value of b for the ^{1}S state as it modifies the contribution of the ^{3}S state to the scattering of thermal neutrons by protons; the effect is, however, much smaller than the experimental uncertainty.

A more accurate approximation has been developed by Ramsey (1948). In table 2 the predictions of these formulae for the ^{1}S state are compared with the results of numerical integration at a series of energies up to 10 MeV. Ramsey's method consistently gives better results than the Born approximation, but neither method is sufficiently accurate for the purposes of the present paper. For states with l greater than zero Ramsey's approximation is reliable at all energies.

TABLE 2. ^{1}S PHASE SHIFTS

energy (MeV)	meson mass	Born's approximation	Ramsey's approximation	numerical integration
0·2495	$225\,m_e$	2°·8	6°·8	11°·2
2·392	$275\,m_e$	15°·2	36°·8	51°·0
2·392	$225\,m_e$	17°·3	38°·6	49°·8
4·2	$225\,m_e$	22°·6	45°·4	52°·7
9·94	$275\,m_e$	27°·6	48°·0	53°·1

The approximate formulae for the phase shift are cumbersome on account of the complicated nature of the Coulomb functions. In the scattering of neutrons by protons only well-known functions occur and the approximations are consequently more tractable. At high energies corresponding proton-proton and neutron-proton phase shifts are equal, and so the latter may be computed instead of the former. The error introduced by adopting this procedure at medium energies is small, especially for states with large angular momenta. At 10 MeV the neutron-proton and proton-proton ^{1}S phase shifts differ by approximately 5 %; the percentage difference will be considerably smaller for the ^{3}P state. The neutron-proton phase shifts have therefore been used for the ^{3}P and ^{1}D states; the contribution of these states to the scattered distribution is relatively small and so a small percentage error is unimportant.

In the case of ^{1}S states the wave function was found by a step-by-step numerical integration. The phase shift was then calculated from the asymptotic form of the wave function by means of a formula given by Breit *et al.* (1936). The error in these phase shifts is considerably less than 1 %.

THE SCATTERING OF SLOW PROTONS BY PROTONS

The term 'slow' in this connexion refers to the energy range in which the nuclear force between the protons gives rise to an S scattered wave only (say, from 0·1 to 5 MeV).

The experimental data on proton-proton scattering at energies below 2·4 MeV accumulated as a result of several series of careful investigations carried out in America between 1936 and 1940. A complete bibliography of the papers relating to these experiments is given by Rosenfeld (1947). The empirical data were analyzed in detail by Breit *et al.* (1936) and by Breit, Thaxton & Eisenbud (1939), and they found that the internal consistency of the data was extremely good. This suggests that the experiments are very accurate and consequently are suitable for testing a theory of nuclear forces. More recently, May & Powell (1947) have investigated the scattering of 4·2 MeV protons by protons using a photographic technique.

In table 3 and in figure 1 the theoretical predictions are compared with the empirical data. The agreement cannot be regarded as satisfactory. If the meson mass is $225m_e$ (or lower), the theory leads to phase shifts which are too large at very low energies and which are too small at higher energies; it is therefore not possible to get agreement with experiment by altering the value of b. On the other hand, a meson mass of $275m_e$ leads to too large phase shifts at all the energies considered, and a tolerable agreement with experiment can be obtained by reducing the interaction constant by 2 %. The broken curve of figure 1, which is acceptably close to the empirical data, is constructed from the numerical integrations of Lubanski & de Jager (1948) who used a meson mass of $274m_e$ and a b-value of 1·526. Agreement with experiment can also be obtained with higher masses of the meson; in fact, the empirical results would favour a slightly higher mass. Hoisington, Share & Breit

TABLE 3. ^{1}S PHASE SHIFTS

energy (MeV)	$m = 225m_e$ $b = 1{\cdot}530$	$m = 275m_e$ $b = 1{\cdot}558$	experimental value
0·2495	11°·2	10°·6	9°·1 ± 0°·5
1·200	40°·7	40°·3	35°·9 ± 0°·4
2·392	49°·8	51°·0	48°·1 ± 0°·4
4·2	52°·7	55°·2	54° ± 2°·5

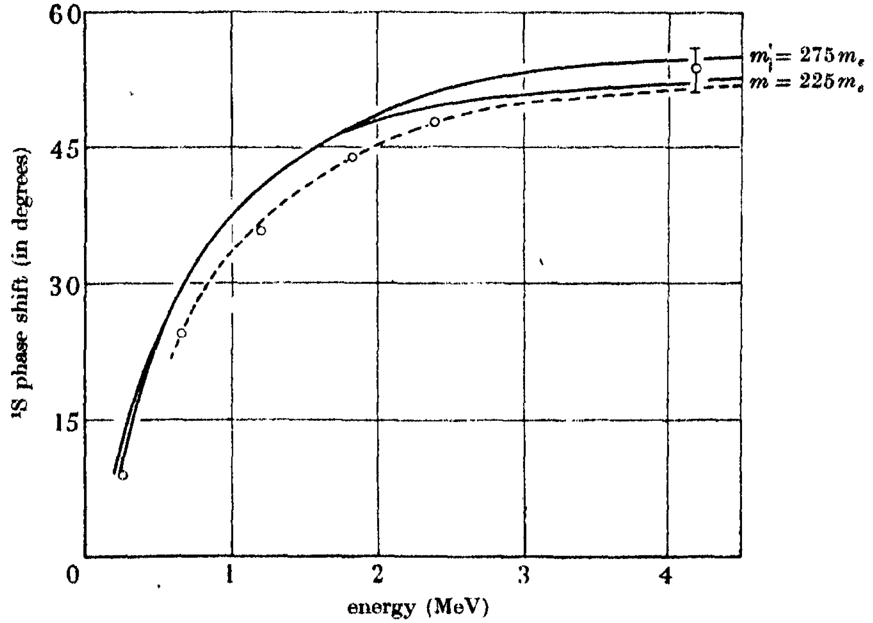


FIGURE 1. The theoretical variation of the ^{1}S phase shift with energy, using the interaction constant determined from the scattering of thermal neutrons by protons; the full curves refer to meson masses of $225m_e$ and $275m_e$. The experimental points are a selection from various authors. The broken curve, which is in good agreement with the empirical data, refers to a meson mass of $274m_e$ and an interaction constant about 2% weaker than that of the neutron-proton interaction.

(1939) found that there was good fit with a meson mass of $326m_e$ and a b-value of 1·547. It is therefore not possible to use the results of the experiments in this energy range to fix the mass of the meson precisely. The value of b used by Hoisington *et al.* (1939) is again 2 % lower than the corresponding value for the neutron-proton interaction, which is about 1·575. The difference in the values of b derived from neutron-proton and proton-proton scattering does not, therefore, disappear when larger meson masses are used.

The difference in the magnitudes of the neutron-proton and proton-proton interactions which is revealed by this analysis hinges on Hanstein's (1941) determination of the cross-section for the scattering of thermal neutrons by protons, which was $21 \pm 1 \times 10^{-24}$ cm.^{2}. This determination limits the value of b to 0·2 % (cf. table 1), while the difference in question is approximately 2 %; it cannot, therefore, be set down to experimental errors. The results of proton-proton scattering correspond to a cross-section of 14×10^{-24} cm.^{2} for the scattering of thermal neutrons by protons. Such a cross-section was found by Simons (1939) and was used by Hoisington *et al.* (1939) to establish the charge independence of nuclear forces for the meson type of interaction of equation (1). Simons's results are, however, less reliable than those of Hanstein, and his experimental arrangement is open to criticism (cf. Rosenfeld 1947).

The experiments in this low energy region do not reveal any P or D scattered waves arising from the nuclear interaction of the protons. From the experimental uncertainties Breit, Thaxton & Eisenbud (1939) have been able to set upper limits for the absolute magnitude of the ^{3}P phase shift at different energies. These in turn may be used to set a lower limit to the mass of the meson. In the Møller-Rosenfeld theory it is found that any meson mass above $270m_e$ is acceptable from this point of view.

THE SCATTERING OF FAST PROTONS BY PROTONS

The technique and equipment which have been used in the experimental investigations of the scattering of fast protons by protons were developed by Wilson & Creutz (1947), who carried out a set of experiments with protons of 8 MeV energy. More recently the apparatus has been used to investigate proton-proton scattering at 10 MeV by Wilson (1947) and at 14·5 MeV by Wilson, Lofgren, Richardson, Wright & Shankland (1947). The phase shifts indicated by these experiments are given in table 4. These have been determined by comparing the empirical results with the general expression containing arbitrary ^{1}S and ^{3}P phases. The experiments at 8 and 14·5 MeV energy are not sufficiently accurate to provide definite evidence of the presence of a P-wave. On the other hand, Wilson's experiments at 10 MeV energy indicate a negative ^{3}P phase shift, which corresponds to a repulsive interaction. Wilson's results are presented in detail in figure 2,* where they are compared with the best fit which can be obtained by an S-wave only (broken curve) and by allowing both S- and P-waves (full curve). Wilson did not measure any absolute cross-section at 10 MeV. However, as the ^{1}S phase shift decreases slowly with energy in this region, the experimental results at 8 and 14·5 MeV may be combined to indicate a ^{1}S phase shift of 52°·7 ± 2° at 10 MeV; Wilson's results have been normalized to this figure at 90° angle of scattering. Peierls & Preston (1947) have pointed out that the ^{3}P phase shift deduced from the empirical data is not sensitive to small changes in the vertical scale of figure 2. The uncertainties in the phase shifts quoted in table 4 have been deduced from the errors estimated by the investigators.

---

* Figure 2 is based on a table of measurements sent to the author by Dr Wilson.

TABLE 4. EXPERIMENTAL PHASE SHIFTS

energy (MeV)	^{1}S state	^{3}P state	investigators
8·0	53° ± 2°	$\vert\eta_1\vert < 0°{\cdot}6$	Wilson & Creutz (1947)
9·94	—	$-0°{\cdot}6 \pm 0°{\cdot}3$	Wilson (1947)
14·5	52° ± 2°	$\vert\eta_1\vert < 3°$	Wilson *et al.* (1947)

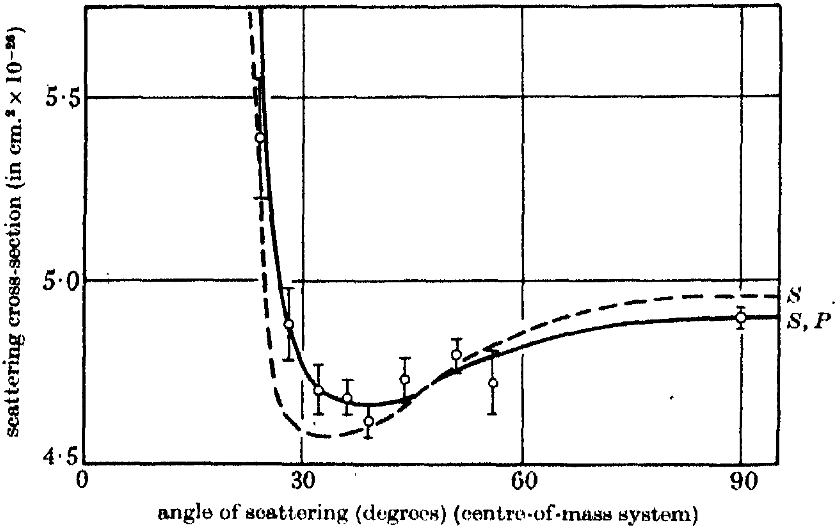


FIGURE 2. The experimental results of Wilson (1947) on the scattering of 10 MeV protons by protons. The broken curve is the best fit which can be obtained if only *S*-scattering is allowed. The full curve gives the best agreement possible with *S*- and *P*-waves; the ^{3}P phase shift is $-0°{\cdot}6$.

Using the values of b indicated by the scattering of slow protons by protons, it is found that the theoretical predictions agree most closely with the empirical ^{1}S phases when the meson mass is assumed to be $270 m_e$. With this meson mass the theory predicts ^{1}S phase shifts of 53°·6, 52°·7 and 51°·3 at the energies 8, 10·and 14·5 MeV respectively. In contrast to the results at low energies, the ^{1}S phase shift is sensitive to changes in the mass of the meson. Using perturbation theory, the phase shift at 10 MeV energy is found to increase by

$$\delta\eta_0 = 0{\cdot}27(\delta m/m)$$

when the mass of the meson is increased by δm; when the meson mass is increased to $300 m_e$, the phase shift is increased by nearly 2°. The uncertainties in the experimental ^{1}S phase shifts in table 4 set limits of $270 \pm 30 m_e$ to the mass of the meson.

The experiments of Wilson (1947) indicate a small negative ^{3}P phase shift at 10 MeV energy (cf. table 4). This is qualitatively in accord with the weakly repulsive interaction predicted by the Møller-Rosenfeld theory for two nucleons in a ^{3}P state. However, the quantitative predictions of this theory are not in very good agreement with the experimental results. The ^{3}P phase shifts expected at 8, 10 and 14·5 MeV energy for several values of the mass of the meson are given in table 5. For meson masses of $275m_e$ and $300m_e$ the magnitude of the phase shift at 10 MeV energy is decidedly larger than the empirical value. To get agreement with experiment a mass in the $350m_e$ to $600m_e$ range must be used. Such a mass could not, however, lead to the observed ^{1}S phase shifts.

TABLE 5. THEORETICAL ^{3}P PHASE SHIFTS

meson mass	b	8 MeV	10 MeV	14·5 MeV
$275m_e$	−0·796	−1°·24	−1°·64	−2°·27
$300m_e$	−0·777	−1°·08	−1°·35	−1°·91
$400m_e$	−0·68	−0°·49	−0°·64	−0°·97
$600m_e$	−0·62	−0°·16	−0°·21	−0°·34

At energies of 8 MeV or higher the Møller-Rosenfeld theory predicts a small positive ^{1}D phase shift, which is about 0°·5 at 10 MeV energy for a meson mass of $275m_e$. The presence of this wave would be difficult to detect experimentally as its contribution to the scattering is similar to that of a small negative ^{3}P phase shift. For this reason its presence exaggerates the discrepancy between theory and experiment mentioned in the previous paragraph; it effectively increases the magnitude of the ^{3}P phase shifts given in table 5.

DISCUSSION

The experiments on the scattering of slow protons by protons indicate a meson mass of $275m_e$ or greater. To get agreement between theory and experiment, however, it has to be assumed that the nuclear interaction between two protons is 2 % weaker than the corresponding interaction between a neutron and a proton. The experimental ^{1}S phase shifts in the 8 to 15 MeV energy range indicate a meson mass of $270 \pm 30m_e$. Thus, a meson mass between $275m_e$ and $300m_e$ is consistent with all the experimental data on the nuclear interaction between two protons in a ^{1}S state. However, such a mass leads to an absolute ^{3}P phase shift at 10 MeV energy which is significantly greater than that observed by Wilson (1947), and a mass of $350m_e$ or greater is required to reduce the theoretical phase shift to the experimental value.

From the results of this investigation it is clear that the Møller-Rosenfeld version of the meson theory of nuclear forces is unsatisfactory and that some important modification must be envisaged. The first difficulty of the theory is that the mass of the meson deduced from the scattering of protons by protons is too large to allow the identification of these particles with the charged particles found in cosmic radiation; the best estimate of the mass of the latter particles is $180m_e$ to $200m_e$. The second difficulty encountered is the difference in magnitude of the neutron-proton and proton-proton interactions. The third difficulty faced by the theory is that the mass of the meson deduced from the results of experiments on the scattering of protons by protons does not agree with the values indicated by other empirical data. From the binding energies of the lightest nuclei Fröhlich, Ramsey & Sneddon (1947) obtained a value of $225 m_e$ for the mass of the meson to be used in the Møller-Rosenfeld theory. Ramsey (1947) has also found that a meson mass of $225 m_e$ is consistent with the results of experiments on the scattering of fast neutrons by protons. This value cannot be reconciled with the range of mass 275 to $300 m_e$ indicated by proton-proton scattering.

The Møller-Rosenfeld theory has met with more success than any other version of the meson theory of nuclear forces. It would therefore appear that modification, rather than complete abandonment, is the proper approach to the difficulties outlined above. Experiments on the scattering of fast neutrons by protons have confirmed that the interaction between two nucleons in a ^{1}P state is strongly repulsive, and the experiments of Wilson (1947) have confirmed that the interaction between two protons in a ^{3}P state is weakly repulsive. These features are important characteristics of the Møller-Rosenfeld theory, and their confirmation implies that the set of eigenvalues of the operator G of equation (1) is either correct or need be modified only slightly. The radial dependence of the static interaction, namely $r^{-1}e^{-\lambda r}$, has a theoretical but no experimental justification. The difficulties which the theory meets when applied to proton-proton scattering suggest that the radial dependence of the interaction has too small a 'curvature'; that is, if the constants of the theory are adjusted to give a good description of the nuclear interaction at small distances, then it will predict too large an interaction at large distances. This means that the effective value of λ, and so the effective meson mass, will increase as the distances between the nucleons is increased. The meson mass indicated by the scattering of protons by protons will then be larger than that for the neutron-proton interaction because of the fact that the protons are kept apart by the electrostatic repulsion. The mass difference should, however, decrease rapidly with energy. This is actually the case. The experimental ^{1}S proton-proton phase shifts in the 8 to 15 MeV energy region correspond to a meson mass of $240 \pm 30 m_e$ if the interaction constant indicated by neutron-proton scattering is adopted; this mass is not significantly different from that found for the neutron-proton interaction. The wrong radial dependence may also be responsible for the apparent difference in the strengths of the neutron-proton and proton-proton interactions; Hoisington *et al.* (1939) have shown that such a difference of the order of a few per cent can arise in this way. The absolute ^{3}P phase shift, for which the theory predicts too large a value, would be decreased if the interaction (1) were replaced by one of greater 'curvature'.

In conclusion the writer would like to thank Dr H. Fröhlich and Professor L. Rosenfeld for the interest they have shown in this work. He is also indebted to Mr de Jager for a preliminary account of the researches of Lubanski & de Jager (1948), and to Miss M. J. Littleton for assistance with the computations.

REFERENCES

Breit, Condon & Present 1936 *Phys. Rev.* **50**, 825.
Breit, Thaxton & Eisenbud 1939 *Phys. Rev.* **55**, 1018.
Fröhlich, Ramsey & Sneddon 1947 *Rpt. Camb. Conference*, vol. 1, 166.
Hanstein 1941 *Phys. Rev.* **59**, 489.
Hoisington, Share & Breit 1939 *Phys. Rev.* **56**, 884.
Hulthén 1944 *K. fysiogr. Sällsk. Lund Förh.* **14**, no. 21.
Lubanski & de Jager 1948 *Physica* (in the Press).
May & Powell 1947 *Proc. Roy. Soc.* A, **190**, 170.
Møller & Rosenfeld 1940 *K. danske vidensk. Selsk. mat.-fys. Medd.* **17**, no. 8.
Peierls & Preston 1947 *Phys. Rev.* **72**, 250.
Ramsey 1947 *Proc. Roy. Soc.* A, **191**, 195.
Ramsey 1948 *Proc. Camb. Phil. Soc.* **44**, 87.
Rosenfeld 1945 *K. danske vidensk. Selsk. mat.-fys. Medd.* **23**, no. 13.
Rosenfeld 1947 *Nuclear forces*. North Holland Publishing Co.
Simons 1939 *Phys. Rev.* **55**, 792.
Wilson 1947 *Phys. Rev.* **71**, 384.
Wilson & Creutz 1947 *Phys. Rev.* **71**, 339.
Wilson, Lofgren, Richardson, Wright & Shankland 1947 *Phys. Rev.* **71**, 560.

Determination of the viscosity of molten salts by a capillary method. The system cadmium chloride-cadmium bromide

BY H. BLOOM, B. S. HARRAP AND E. HEYMANN
Chemistry Department, University of Melbourne

(*Communicated by Sir David Rivett, F.R.S.—Received* 13 *December* 1947)

The knowledge of the viscosity of molten salts and their mixtures is important (1) for testing recent theories by Eyring and by Frenkel, and (2) for confirming or modifying conclusions regarding the constitution of certain molten salt mixtures which were investigated by Bloom & Heymann. A capillary method based on the Ubbelohde-Bingham principle has been adapted to the conditions at high temperature. The main feature of the method is an automatic electric recording device, involving a balanced Wheatstone net with amplifier and an electric clock.

The isotherms of viscosity plotted against molar fraction in the system $CdCl_2$-$CdBr_2$ are linear. The activation energy of viscosity is 1·6 to 1·8 times as great as the activation energy of ionic migration.

INTRODUCTION

Eyring and his colleagues (Eyring 1936; Ewell & Eyring 1937; Glasstone, Laidler & Eyring 1941) have applied the theory of absolute reaction rates, combined with the 'hole' theory of liquids, to the problem of viscosity. They consider the holes in a liquid as playing a part similar to that played by the molecules in a gas. Their investigations have yielded interesting material concerning the total and free energies of activation of viscosity, their relation to the energy of vaporization, and the possible causes of their variation with temperature in some systems. The latter have also been discussed by Ward (1937). These concepts have been mainly applied to molecular liquids. An earlier theory by Frenkel (1926, 1946), based on the connexion between viscosity and self-diffusion, has also been applied to ionic liquids. Both schools have interpreted equations relating viscosity with the composition of a mixture, and Frenkel (1946) has proposed and discussed various forms of mixture laws of viscosity and electric conductivity of molten salts.

Experimental data on ionic liquids (molten salts) are scarce. It is the object of our investigations to supply the experimental material for the application of the concepts mentioned above to molten salts.

Bloom & Heymann (1947), who discussed the activation energies of viscosity and of ionic migration of molten salts, found that the former are always greater than the latter. Thus the energy barriers involved in viscous flow are different from those involved in ionic migration. This is probably due to the fact that, whilst conductivity is mainly determined by the ionic species which possesses the higher mobility (usually the cations), viscosity depends mainly on the ionic species which possesses the smaller mobility (usually the anions) (Frenkel 1946; Bloom & Heymann 1947). Bloom & Heymann have also investigated the isotherms of conductivity of a number of binary salt mixtures. These give an interesting insight into the possible constitution of these mixtures (e.g. presence of complex ions), but a parallel investigation of the viscosity of such mixtures is desirable.

EXPERIMENTAL

For liquids of low viscosity, the most frequently used methods of measurement are the capillary method, the oscillating sphere or disk method, and the method of stationary deviations. The last two methods involve considerable constructional and operational difficulties at high temperatures, and the oscillating sphere method requires a large quantity of platinum. To illustrate such operational difficulties, it may be mentioned that Lorenz & Kalmus (1907), who used the latter method, had to discard 30 % of their measurements because the decrements were not logarithmic. We therefore decided to adapt the capillary method to our use.

In the capillary method, the coefficient of viscosity (η) is calculated from the rate of laminar flow by the well-known equation of Poiseuille. Considerable divergence of opinion prevails regarding the conditions in which the Hagenbach (kinetic energy) correction is applicable. Grüneisen (1905), in an exhaustive investigation on the range of validity of Poiseuille's equation, found that no correction was justified in instruments in which the capillary had a trumpet-shaped end leading to a large reservoir, and concordant results were obtained for the product of the driving pressure (p) and the time of flow (t) of a given volume of liquid at different pressures. It can easily be seen that this is a criterion for the validity of Poiseuille's equation without correction. We therefore constructed our viscometers with trumpet-shaped ends, in a fairly large reservoir, and calibrated them empirically in a molten salt, the viscosity of which had been determined by independent methods. The constancy of $p \times t$ with varying pressure in the experimental range employed is shown in table 1 for a mixture containing 55·0 mole % $CdCl_2$ and 45·0 mole % NaCl at 411° C.

The onset of turbulence was observed in a number of experiments at high temperatures. Values for the Reynolds number calculated from such experiments were of the order of 1500, suggesting that the shape of the capillary (see below) was such that it did not give rise to local turbulence. The constancy of $p \times t$ was tested in all experiments, and only such experiments in which this criterion was fulfilled were used for a calculation of the viscosity.

TABLE 1

p (cm. Hg)	t (sec.)	$p \times t$
4·35	180·8	786
8·40	92·5	777
13·41	58·5	784
16·54	47·3	781

The choice of the glass for viscometer and container depended on the melting-points of the system under investigation. Moncrieff, Supremax (Jena) and B.T.H. glass (C14 and C46) were used, B.T.H. being particularly suitable because of its high softening point (C14: 825 to 850° C; C46: 875 to 900° C) and its resistance to chemical attack by the molten salt. Silica is unsuitable because it is strongly attacked by many molten salts.

The apparatus (figure 1) consisted of a capillary C (of 12 to 13 cm. length) with trumpet-shaped ends; the diameter varied with the viscosity of the system under investigation and was of the order of 0·5 mm. The upper end of the capillary was sealed to a glass reservoir R, of 6 cm. length and 1·4 cm. internal diameter. The reservoir was sealed to a tube D, 36 cm. long and 1·4 cm. wide, and there was a constriction N between reservoir and tube, and a side arm S which connected with a manostat. The container was 55 cm. long and of internal diameter of about 3 cm.

Since it is difficult to observe visually the time taken for the melt to flow past two marks, we used an electric recording device. Three platinum contacts were sealed through the Moncrieff glass tube T. When the melt passed the lower contacts at B, equivalent to the lower mark of the viscometer, a current was caused to flow through the two arms of a Wheatstone net which had previously been balanced (figure 2). This current started an electric clock. When the melt reached the upper contact E, equivalent to the upper mark of the viscometer, the Wheatstone net was balanced again and the current ceased to flow through the bridge, thus stopping the clock. The current supplied to the bridge was transformed from the electric mains supply (250 V) down to 16 V. A valve amplifier, designed and constructed by A. E. Ferguson,* was placed in the circuit between the bridge and the clock as shown in figure 2, for the following reason. The alternating current of the mains supply is not of pure sine form and causes a certain amount of electrolysis of the melt between the wires A and B which is troublesome. By placing the amplifier in the bridge circuit with the clock, the current and thus the amount of electrolysis is effectively reduced. The clock used was a Telechron self-starting clock.

A buzzer was also included in the circuit as shown. This was in order to follow the expulsion of the melt from the viscometer without having to use the clock.

---

* Tribophysics Section of the Council for Scientific and Industrial Research of Australia; now of Department of Electrical Engineering, University of Melbourne.

The pressure to drive the melt through the viscometer was provided by using two 20 l. carboys as a vacuum reservoir. By means of this large reservoir, the pressure could be maintained constant to within 0·02 cm. of mercury during each experiment. The expulsion of the melt from the viscometer was effected by compressed nitrogen.

The furnace, which permitted observation of the melt during the experiment, was similar to that described by Bloom & Heymann (1947). The temperature was measured by two thermocouples and a sensitive Leeds and Northrup potentiometer. The hot junctions were arranged at the upper and lower 'marks' of the viscometer. Before any reading was recorded, the temperature was constant for about 10 min., and the temperature at different levels did not vary by more than 1° C.

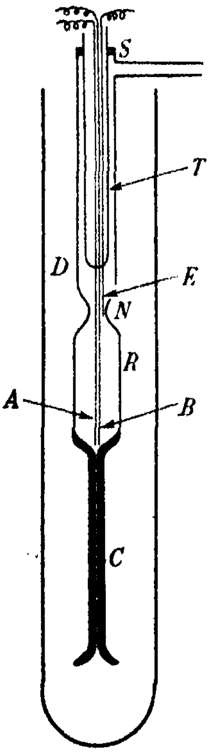


Figure 1. Viscometer.

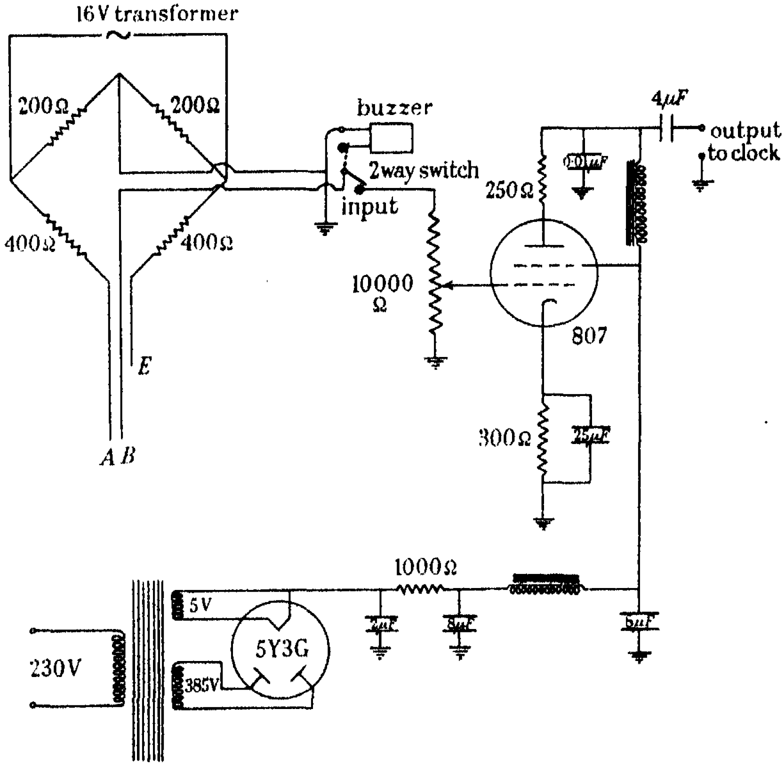


Figure 2. Electrical recording device.

Before the actual determination of viscosity, the positions of the melt, viscometer and thermocouples were adjusted so that no correction had to be made to the pressure read on the mercury manometer, due to the hydrostatic head produced by the different liquid levels inside and outside the viscometer bulb. The dimensions of the bulb and the container were chosen, and the various levels of liquid arranged so that the levels of liquid both inside and outside the bulb were just crossing at half-time. In this way, the head to be added to the driving pressure in the first half of the experiment is balanced by that to be subtracted in the second half (Ubbelohde 1908; Bingham & Jackson 1917, principle).

Molten potassium nitrate was used for calibration; the viscosity values used were those of Dantuma (1928), whose experimental procedure appeared to be more satisfactory than that of most other workers. The temperature coefficient of viscosity of potassium nitrate determined by our method was the same as that observed by Dantuma. The viscometers were recalibrated at frequent intervals to ascertain that the constants had not changed as a consequence of the attack of the molten salt. No change greater than 0·3 % was recorded during the lifetime of any viscometer.

Experimental accuracy

The temperature was constant within 1° C, corresponding to a maximum error of 0·5 % in viscosity. The time was measurable within 0·5 sec. by the electric clock, the times involved in the experiment usually being 50 to 100 sec. Thus the accuracy of the determinations of time was 0·5 to 1·0 %. The driving pressure remained constant within 0·2 % during an experiment. It can be shown easily that the errors introduced by the expansion of the viscometer and of the platinum are negligible. The maximum errors of the determination of viscosity would thus be of the order of 1·0 to 1·5 %. Actually the reproducibility of most experiments was of this order.

RESULTS

The performance of the method was tested in the system cadmium chloride-cadmium bromide,* where our previous work on electric conductivity led us to expect viscosity isotherms of a simple shape. The viscosities of pure cadmium chloride and cadmium bromide are shown over a range of temperatures in figure 3, together with those of cadmium chloride determined by Karpachev & Stromberg (1938). The values that these authors obtained by the oscillating sphere method are about 2 % higher than ours.

Figure 4 shows two isotherms of viscosity plotted against the molar fraction of $CdCl_2$ in the mixture. The isotherms are straight lines. This is probably one of very few cases in the field of liquid mixtures in which the viscosity is a linear function of the molar fraction. With mixtures of non-ionic liquids, even if the systems are ideal with respect to fugacity or any other property connected with the thermodynamic potential, the viscosity isotherms indicate deviations from additivity (cf. Hatschek 1928).

For many molten salts the viscosity may be represented by the equation

$$\eta = \text{const.} \times e^{B/RT},$$

where B is the activation energy characteristic for the configurational changes involved in viscous flow, whilst the electric conductivity, according to Bloom & Heymann (1947), can be represented by a similar equation

$$\kappa = \text{const.} \times e^{-C/RT},$$

where C is the corresponding activation energy characterizing ionic migration. Nevertheless, exceptions have been noted in several systems.

---

* For purification and analysis, see Bloom & Heymann (1947).

Thus, whilst in very dilute aqueous solutions, a relation

$$\kappa \times \eta = \text{const.}$$

has often been found to hold approximately (Walden 1906), for molten salts we have a relation

$$\kappa^m \times \eta = \text{const.},$$

where $m = B/C > 1.$

According to Bloom & Heymann (1947), m is between 3 and 7 for univalent chlorides, and about 2 for bivalent chlorides, showing that the energy barriers involved in

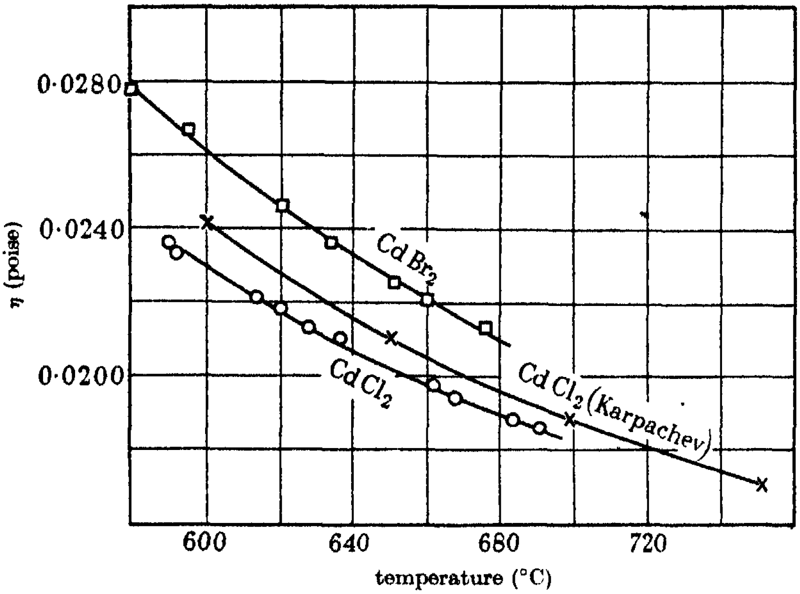


FIGURE 3. Plot of viscosity of pure $CdCl_2$ and pure $CdBr_2$ against temperature.

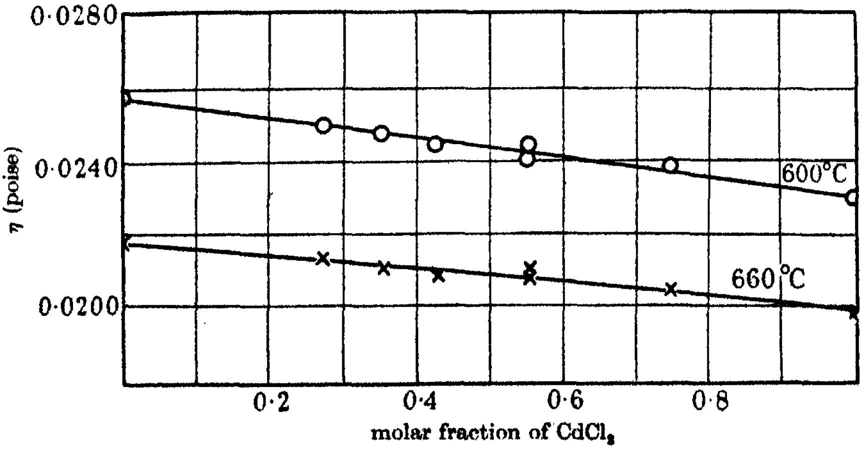


FIGURE 4. Plot of viscosity of $CdCl_2$-$CdBr_2$ mixtures against molar fraction of $CdCl_2$ (600 and 660° C).

viscous flow are greater than those involved in ionic migration. A similar value for m is found in the system $CdCl_2$-$CdBr_2$ (table 2).

TABLE 2

Mole % $CdCl_2$	0	27·2	35·6	41·7	55·5	74·8	100
B (in kcal. mole^{-1})	4·5	4·4	4·4	4·3	4·3	4·3	4·0
C (in kcal. mole^{-1})	2·8	2·7	—	2·5	2·4	2·3	2·3
m	1·6	1·6	—	1·7	1·8	1·8	1·7

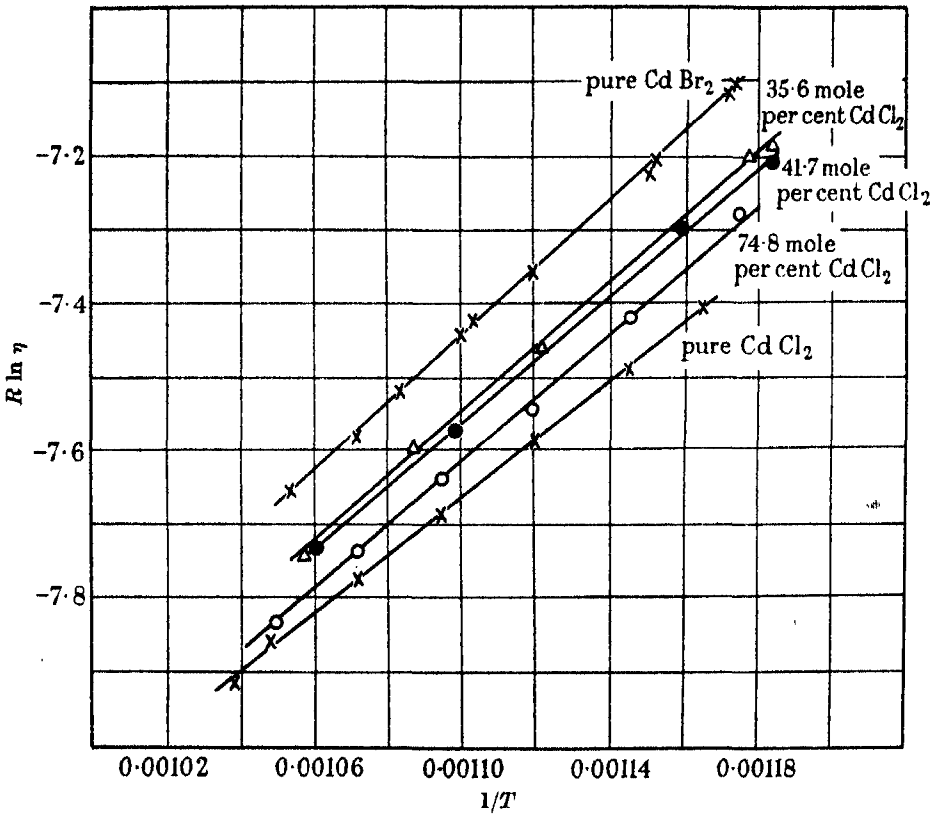


FIGURE 5. Plot of $R \ln \eta$ against $1/T$ for a number of mixtures of $CdCl_2$ and $CdBr_2$. (Pure $CdCl_2$; 74·8 mole % $CdCl_2$; 41·7 mole % $CdCl_2$; 35·6 mole % $CdCl_2$; pure $CdBr_2$.)

B does not vary appreciably with temperature, the plot of $R \ln \eta$ against $1/T$ being linear in the temperature interval under investigation (figure 5).

The theoretical significance of these results will be discussed in conjunction with other systems in a future publication.

REFERENCES

Bingham & Jackson 1917 *Sci. Pap. Bur. Stand., Wash.*, no. 298.
Bloom & Heymann 1947 *Proc. Roy. Soc.* A, **188**, 392.
Dantuma 1928 *Z. anorg. Chem.* **175**, 1.
Ewell & Eyring 1937 *J. Chem. Phys.* **5**, 726.

Eyring 1936 *J. Chem. Phys.* **4**, 283.
Frenkel 1926 *Z. Phys.* **35**, 652.
Frenkel 1946 *Kinetic theory of liquids.* Oxford: Clarendon Press.
Glasstone, Laidler & Eyring 1941 *The theory of rate processes.* New York and London: McGraw-Hill Book Co.
Grüneisen 1905 *Wiss. Abh. phys.-tech. Reichsanst., Berl.*, **4**, 151.
Hatschek 1928 *The viscosity of liquids.* London: Bell and Sons.
Karpachev & Stromberg 1938 *J. Phys. Chem. U.S.S.R.* **11**, 852.
Lorenz & Kalmus 1907 *Z. Phys. Chem.* **59**, 244.
Ubbelohde 1908 *Handb. Chem. u. Techn. Öle*, **1**, 340. (Cf. Hatschek 1928.)
Walden 1906 *Z. Phys. Chem.* **55**, 246.
Ward 1937 *Trans. Faraday Soc.* **33**, 88.

A general kinetic theory of liquids

V. Liquid He II

By H. S. Green, *Edinburgh University*

(*Communicated by M. Born, F.R.S.—Received* 20 *December* 1947)

This paper fulfils the programme announced in part IV of the series by examining the detailed application of the general quantum theory of fluids to liquid He II. It is shown that although the classical laws of hydrodynamics and thermodynamics are formally obeyed by the quantum liquid, important divergences arise in the manner of their interpretation. Without introducing any assumption other than the validity of the general laws of quantum mechanics, it is possible to explain qualitatively and quantitatively the well-known properties of liquid helium, including the thermodynamic discontinuity, the thermomechanical effect, the transfer effect, and the presence of thermal waves in the liquid. The connexion with earlier theories is traced, and it is shown that although they are in agreement with the theory here proposed in certain respects, their fundamental concepts are in need of considerable revision. The theory can be applied with slight changes to the phenomenon of superconductivity.

1. INTRODUCTION

The classical theory of liquids developed in the first three papers of this series (Born & Green 1946; Green 1947; Born & Green 1947*a*) and the quantization effected in the subsequent paper (Born & Green 1947*b*) led to a conception of the nature of viscous and thermal conductive processes in liquids which the authors hoped would be of assistance in the understanding not only of normal liquids, but also of the 'quantum liquids' of which liquid He II is the simplest example. (The electrons in certain metals at low temperatures may be cited as a second example.)

Many distinguished contributions have been made to the experimental evidence relating to the phenomena which are studied in this paper. Among them, those of Kapitza (1941*b*) on the thermomechanical effect, and of Daunt & Mendelssohn (1939) on the related transfer effect appear to be of special importance, as they demonstrate most clearly the existence of a special kind of reversible process which can occur in thin layers of the liquid. Other non-equilibrium phenomena are readily interpreted as manifestations of the same type of reversible process, adulterated in many instances by the simultaneous operation of irreversible effects. The remarkable feature exhibited by all these experiments is the apparent failure in He II of two well-established physical laws: the law of thermal conduction, which requires the heat flow to be proportional to the temperature gradient, and the law of viscosity, which requires the velocity gradient to be proportional to the non-hydrostatic pressure forces.

No satisfactory explanation has yet been given of these phenomena. They lie outside the province of ordinary reversible thermodynamics, which requires for its validity that changes of any kind should occur with infinite slowness. Further, no account can be expected of the anomalous viscosity and thermal conduction in He II without prior understanding of the corresponding processes in ordinary liquids. In spite of these handicaps, several attempts have been made to elucidate certain aspects of the problem, of which the best known are due to London (1938, 1939), Tisza (1938, 1940, 1947), and Landau (1941, 1944). The first two of these authors endeavoured to connect superfluidity with the 'condensation' which occurs in a gas of non-interacting Bose particles at a temperature not far from the λ-point. They were not able, however, to take adequate account of the interaction between the helium molecules; moreover, it seemed improbable that Bose-Einstein statistics could be of crucial importance for the existence of superfluidity, since entirely analogous phenomena were known to occur in superconductors, where the electrons obey Fermi-Dirac statistics. Landau, while retaining the idea of London and Tisza that He II could be regarded as a mixture of two phases of which only one could participate in the superfluid motion, eliminated these objections by introducing a new description of the superfluid as a liquid devoid of 'rotons', rather vaguely defined as states of quantized vortex motion. The theoretical foundations of Landau's work were very insubstantial; nevertheless, his theory was in close relation to the experimental evidence, and justified itself by predicting the existence of two kinds of wave propagation for He II and therefore two velocities of 'sound'.

The present paper reveals that the limited successes of the earlier theories are derived principally from the two-phase conception of the quantum liquid; yet it appears clearly that this is an over-simplification of the real situation, which demands a different physical interpretation. In fact, the classical laws of mechanics and thermodynamics hold only in a modified form, in which the difference π between the thermodynamic and 'kinetic' pressures plays an important part. While many details remain to be investigated, the author believes that all the properties of He II can be explained by the theory developed below in a way not only satisfactory to the physical intuition, but requiring no assumption other than the accepted laws of quantum mechanics. Incidentally, the properties of a quantum liquid are derived in their most general form, which can easily be applied to the electrons in a metal and the phenomenon of superconductivity; one requires only to replace the ordinary pressure tensor by the electromagnetic stress tensor, and to represent the almost stationary metallic ions by a conservative field of force.

2. HE II IN EQUILIBRIUM

Since the radial distribution function plays an important part in determining the properties of both classical and quantum liquids, it is desirable, as a preliminary step, to ascertain the general properties of this function in liquid helium at very low temperatures where the usual methods of statistical mechanics (cf. de Boer & Michels 1939) cannot be applied. The method here adopted was indicated in § 4 of part IV, where it was shown that the density matrix $\rho_2(\mathbf{x},\mathbf{x}')$ relating to a pair of particles may be expanded in the form

$$\rho_2(\mathbf{x},\mathbf{x}') = \sum_{\lambda} a(\lambda)\,\psi_2(\lambda,\mathbf{x})\,\psi_2^*(\lambda,\mathbf{x}'), \tag{2·1}$$

where $a(\lambda)$ is the probability of finding a pair of molecules in the unperturbed eigenstate λ, characterized by the eigenfunction $\psi_2(\lambda,\mathbf{x})$ which is the solution of the equation

$$\left\{\frac{1}{2m}(\mathbf{p}^{(1)2}+\mathbf{p}^{(2)2})+\phi(r)\right\}\psi_2(\lambda,\mathbf{x}) = E'\psi_2(\lambda,\mathbf{x}). \tag{2·2}$$

The λ comprise the quantities $\mathbf{p} = \mathbf{p}^{(1)}+\mathbf{p}^{(2)}$, m_3, where $\mathbf{m} = \frac{1}{2}\mathbf{r}\times(\mathbf{p}^{(2)}-\mathbf{p}^{(1)})$, l, satisfying $l(l+1)\hbar^2 = \mathbf{m}^2$, and $E = \frac{1}{4m}(\mathbf{p}^{(2)}-\mathbf{p}^{(1)})^2+\phi(r) = E'-\frac{1}{4m}\mathbf{p}^2$. Then

$$\psi_2(\lambda,\mathbf{x}) = e^{i\mathbf{p}\cdot\mathbf{x}/\hbar}\,Y_l(\theta,\phi)\,r^{-1}\psi(r), \tag{2·3}$$

where Y_l is a spherical harmonic of order l, and $\psi(r)$ satisfies

$$\left[\frac{d^2}{dr^2}-\frac{l(l+1)}{r^2}+\frac{m}{\hbar^2}\{E-\phi(r)\}\right]\psi(r) = 0. \tag{2·4}$$

The radial distribution function is derived from $\rho_2(\mathbf{x},\mathbf{x}')$ simply by putting $\mathbf{x} = \mathbf{x}'$, so that

$$n_2(\mathbf{x}) = \sum_{\lambda} a(\lambda)\,|\,\psi_2(\lambda,\mathbf{x})\,|^2. \tag{2·5}$$

As was shown in part IV, if the perturbation energy V_2 experienced by the given pair of molecules on account of the other molecules in the liquid is assumed to be Hermitian, the $a(\lambda)$ occurring in (2·1) and (2·5) will have the same value in equilibrium for all states λ between which transitions are not forbidden by the vanishing of the matrix elements of V_2. Now the effect of the other molecules, reflected in the form of the matrix V_2, is to make n_2 tend to a constant value c, equal to r_1^2 in the infinite fluid, when r is large. Thus

$$n_2(\mathbf{x})\to c \quad \text{as} \quad r\to\infty. \tag{2·6}$$

Subject only to this condition, one will have, in equilibrium,

$$a(\lambda) = e^{-E/kT}. \tag{2·7}$$

From this it is apparent that the behaviour of each individual term $a(E)\,|\,\psi_2(E,\mathbf{x})\,|^2$ in the expansion (2·5) depends strongly on E at low temperatures, and, as a result, near absolute zero, one may write

$$n_2(\mathbf{x}) = c\,|\,\psi_2(E_0,\mathbf{x})\,|^2, \tag{2·8}$$

where E_0 is the lowest fully occupied energy level. Now, for $E_0 < 0$ or $E_0 > 0$, $r^{-1}\psi(r)$ tends to zero or infinity for large r, so that the boundary condition (2·6) is violated. Hence $E_0 = 0$, and in the normal condition all molecule-pairs are in the state of zero energy at absolute zero.

This result does not affect the possibility of a limited occupation of the negative energy states. (In the instance of helium, on account of the weakness of the attraction between the molecules, it is unlikely that there are more than one.) In particular, near the walls at the boundary of the fluid, the condition (2·6) no longer holds, and one may substitute the ordinary Schrödinger boundary condition, which is satisfied only in the negative energy states. Hence one is led to predict the existence of liquid helium in an abnormal, negative energy state near the walls of the containing vessel; this is confirmed in the exact treatment of non-uniform conditions developed in the following sections.

To determine $n_2(r)$ for the normal state, one need only solve (2·4) for positive values of E; the Lennard-Jones formula for $\phi(r)$ is chosen because of the hardness of the repulsive term νr^{-12}, which allows one to consider only the attractive term $-\mu r^{-6}$, provided that the part of the wave function corresponding to distances below the potential minimum is subsequently ignored. With the further substitutions $\chi = r^{-\frac{1}{2}}\psi$ and $t = r^{-2}$, (2·4) reduces to

$$\frac{d^2\chi}{dt^2} + \left\{\frac{mE}{4\hbar^2t^3} - \left(\frac{l}{2} - \frac{1}{4}\right)\left(\frac{l}{2} + \frac{3}{4}\right)t^{-2} + \frac{m\mu}{4\hbar^2}\right\}\chi = 0. \tag{2·9}$$

When $E = 0$, this equation has the solution

$$\chi = t^{\frac{1}{2}}J_{(\frac{1}{2}l-\frac{1}{4})}(\beta t), \quad \beta = \left(\frac{m\mu}{4\hbar^2}\right)^{\frac{1}{2}}, \tag{2·10}$$

from which it is clear that only the value $l = 0$ is consistent with the boundary conditions. The form of the radial distribution function obtained from the solution (2·10) is shown in figure 1, and is very different from that characteristic of classical liquids, having a minimum instead of the usual maximum near the minimum of $\phi(r)$. The variation of $n_2(r)$ with temperature can be taken into account by making use of the solution of (2·9) with positive values of E, which is obviously obtained by writing $t - (2mE/3\hbar^2)$ for t in (2·10), when E is small, and making use of the formula

$$n_2(r) = r^{-2}\int_0^\infty e^{-E/kT}\,|\,\psi_2(E,r)\,|^2\,dE. \tag{2·11}$$

The resulting curve, within its range of validity, does not differ significantly from the curve shown in figure 1.

It is clear from (2·9) that the form of the radial distribution function will undergo a considerable modification when energies of the order of $2\hbar^2/mr_1^2$ ($r_1 = 3$ A) are excited, due to contributions from the states $l = 2, 4,$ etc. This corresponds to a temperature of $2\hbar^2/mkr_1 \sim 2°$ K, near the λ-point. It is known that the effect of the higher angular momentum states is to contribute to the normal maxima of the radial distribution function, so that one is led to expect that the unusual features of liquid helium will disappear above 2° K. These plausibility considerations are confirmed by the more precise calculations of the next section, but they are of interest here because they tend to confirm the suggestion of Landau (1941) that the unusual features of He II are associated with a deficiency of 'rotons', or quantized rotational states.

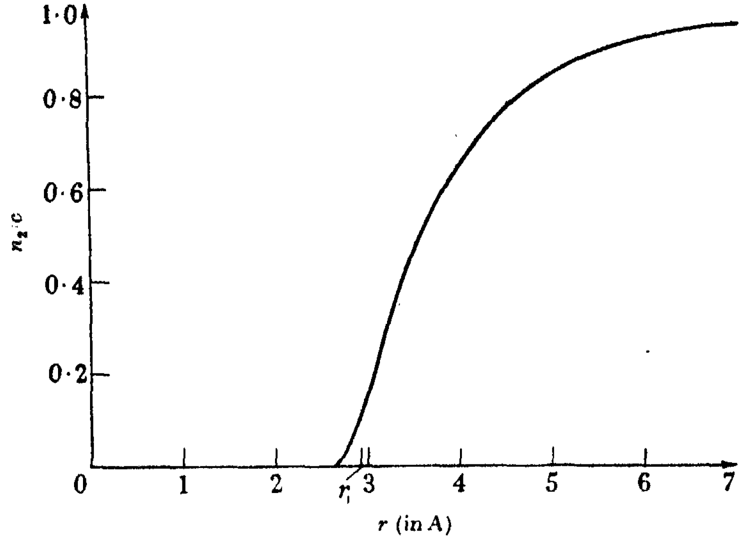


FIGURE 1. Radial distribution function for liquid helium in the normal state near absolute zero. *Distance of potential maximum* $r_1 = 2{\cdot}93$ A.

3. THERMOMECHANICS OF He II

In part IV certain definite conclusions were reached concerning the similarities and differences which may be expected between an ordinary liquid and the 'quantum' modification which must occur, in the absence of solidification, at low temperatures. These conclusions may be briefly summarized as follows.

The equation of motion
$$mn\frac{d}{dt}\mathbf{u}+\frac{\partial}{\partial \mathbf{x}}\cdot\mathsf{p}_1 = n\mathbf{F}, \tag{3·1}$$

and the law of viscosity
$$\mathsf{p}_1' = -2\mu\,\overgroup{\frac{\partial}{\partial \mathbf{x}}\mathbf{u}} \tag{3·2}$$

for the non-hydrostatic part of the pressure tensor, remain formally correct, but the 'kinetic' pressure tensor p_1 differs, even in equilibrium, from the thermodynamic pressure tensor p which is directly measurable. Similarly, the equation of energy transfer
$$n\frac{dU}{dt}+\frac{\partial}{\partial \mathbf{x}}.\mathbf{q}+\mathsf{k}_1:\left(\frac{\partial}{\partial \mathbf{x}}\mathbf{u}\right) = 0 \tag{3·3}$$

and the law of thermal conduction
$$\mathbf{q} = -\lambda\frac{\partial T_1}{\partial \mathbf{x}}, \tag{3·4}$$

where nkT_1 is the diagonal element of the thermal part k_1 of the kinetic pressure tensor, require no modification, but the 'kinetic' temperature T_1 differs from the thermodynamic temperature T.

Part of the remaining task is therefore the examination of the connexion between the kinetic and thermodynamic quantities. In equilibrium one has, according to § 3 of part IV,

$$p = p_1 + p_3 + p_5 + \ldots, \tag{3·5}$$

where

$$p_j = \frac{kT}{3\iota\hbar VN!} X\left\{\frac{\rho_N \mu_j}{j!(kT)^j}\right\} \tag{3·6}$$

and

$$\mu_{j+1} = \mu_j W_N - W_N \mu_j, \quad \mu_1 = \iota\hbar \sum_{i=1}^{N} \left\{\frac{\mathbf{p}^{(i)2}}{m} - \frac{1}{2}\sum_{j=1}^{N} r^{(ij)}\phi'^{(ij)}\right\}, \tag{3·7}$$

so that

$$\left.\begin{aligned} \mu_2 &= \frac{\hbar^2}{m}\sum_{i,j=1}^{N}\left\{\mathbf{p}^{(ij)}, \frac{\partial\psi^{(ij)}}{\partial\mathbf{x}^{(i)}}\right\}, \quad \psi^{(ij)} = \phi^{(ij)} + \tfrac{1}{2}r^{(ij)}\phi'^{(ij)}, \quad \mathbf{p}^{(ij)} = \mathbf{p}^{(i)} - \mathbf{p}^{(j)}, \\ \mu_3 &= \frac{\iota\hbar^3}{m^2}\sum_{i,j=1}^{N}\left\{\mathbf{p}^{(ij)}\cdot\left\{\mathbf{p}^{(ij)}, \frac{\partial^2\psi^{(ij)}}{\partial\mathbf{x}^{(i)}\partial\mathbf{x}^{(j)}}\right\}\right\} - \frac{2\iota\hbar^3}{m}\sum_{i,j,k=1}^{N}\frac{\partial\phi^{(ij)}}{\partial\mathbf{x}^{(i)}}\cdot\frac{\partial\psi^{(ik)}}{\partial\mathbf{x}^{(i)}}. \end{aligned}\right\} \tag{3·8}$$

Hence

$$p_3 = \frac{1}{6m}\left(\frac{\hbar}{kT}\right)^2\left\{\int n_2 kT_2\left(\tfrac{2}{3}\psi'' + \frac{4}{3r}\psi'\right)d\mathbf{r} - 2\iint n_3 \frac{\phi'(r)}{r}\frac{\psi'(s)}{s}\mathbf{r}.\mathbf{s}\,d\mathbf{r}\,d\mathbf{s}\right\}. \tag{3·9}$$

One can now estimate roughly the critical temperature, which will be identified provisionally with the λ-point in liquid helium, below which the quantum correction $p_3 + p_5 + \ldots$ becomes comparable with p_1, given by

$$p_1 = nkT_1 - \frac{1}{6}\int n_2\phi'(r)\,r\,d\mathbf{r}. \tag{3·10}$$

This is done by substituting the approximation T_1 for T_2 in the first term of (3·9), and by applying a Kirkwood-like approximation to the second term in this equation, which then reduces to

$$p_3 = \frac{1}{6m}\left(\frac{\hbar}{kT}\right)^2\left\{nkT_1\int\frac{n_2}{n}\left(\tfrac{2}{3}\psi'' + \frac{4}{3r}\psi'\right)d\mathbf{r} - \int n_2\phi'(r)\,r\,d\mathbf{r}\int\frac{n_2}{n}\psi'(s)\,s^{-1}d\mathbf{s}\right\}. \tag{3·11}$$

Both the integrals involving ψ in (3·11) have values of the magnitude of

$$-2r_1^{-2}\int\frac{n_2}{n}\phi'(r)\,r\,d\mathbf{r},$$

so that p_3 becomes of the same order as p_1 at a temperature given by

$$(kT)^2 \sim \frac{\hbar^2}{3mr_1^2}\int\frac{n_2}{n}\phi'(r)\,r\,d\mathbf{r}, \tag{3·12}$$

where r_1 is the maximum of the integrand. Using (3·10),

$$T = T_0\left(1 - \frac{p_1}{nkT_1}\right), \tag{3·13}$$

where $T_0 = 2\hbar^2/mkr_1^2 \sim 2°\,\mathrm{K}$ for helium, with $r_1 = 3\,\mathrm{A}$.

The formula (3·13) gives the λ-point, together with its variation with pressure, sufficiently well, considering the roughness of the estimation, to support the expectation that a rigorous evaluation of the series (3·5) will yield complete information concerning the thermodynamic discontinuity at the λ-point, the sharpness of the transition being due to the cumulative effect of the higher terms in the infinite series. For the present, write

$$p_1 - p = \pi, \tag{3·14}$$

so that π almost vanishes above the λ-point and becomes positive just below it, as may be seen from the fact that p_3 is negative. It will be found shortly that values of π can be inferred from a variety of experimental determinations; at the saturation vapour pressure (cf. figure 2), it increases to a maximum at 1·5° K, and becomes zero again at absolute zero.

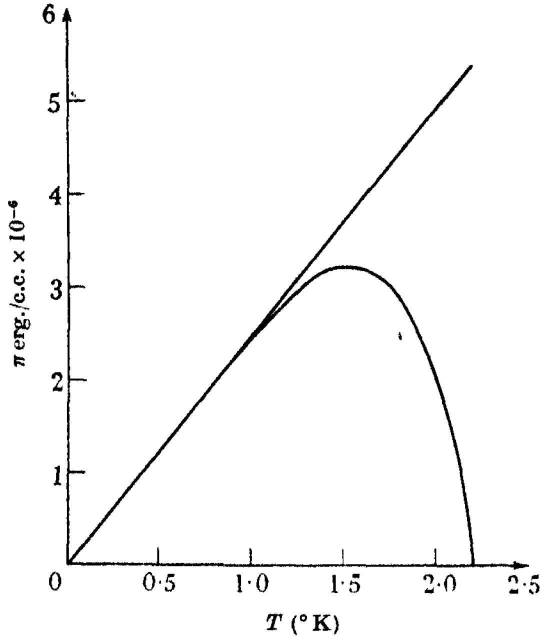


FIGURE 2. π as a function of temperature at the saturation vapour pressure. Upper curve: ground state. Lower curve: normal state. The difference between the two curves is the quantity $nR \doteqdot nTS/m$ measured by Kapitza (1941 *b*, figure 5).

Equation (3·1) may be written in the form

$$mn\frac{d\mathbf{u}}{dt} + \frac{\partial \mathsf{p}}{\partial \mathbf{x}} = -\frac{\partial \pi}{\partial \mathbf{x}} - \frac{\partial}{\partial \mathbf{x}} \cdot \mathsf{p}_1' + n\mathbf{F}, \tag{3·15}$$

which shows that differences in π, such as may occur in the presence of a temperature gradient, will have the effect of obscuring the viscosity and simulating the action of external forces on the liquid; this already secures a qualitative understanding of the thermomechanical phenomena observed in He II.

The thermodynamic pressure tensor p has as yet been given no strict definition applicable to non-uniform conditions. It is now defined by the property that, if $\mathbf{dS}$ is a surface element moving with the liquid, the work done in an elementary displacement $\mathbf{dx}$ of $\mathbf{dS}$ should be $\mathbf{dx} \,.\, \mathsf{p} \,.\, \mathbf{dS}$. This definition evidently coincides with that available in equilibrium conditions, as the negative derivative of the free energy with respect to volume. Consider now a small volume moving with the liquid; the equation of conservation of energy within this volume is

$$\int n\frac{d}{dt}(\tfrac{1}{2}mu^2+U)\,d\mathbf{x}=\int n\frac{dQ}{dt}\,d\mathbf{x}-\int \mathbf{u}\,.\,\mathsf{p}\,.\,d\mathbf{S}, \tag{3·16}$$

where dQ/dt is the rate of absorption of heat energy per molecule, and the external force is omitted from consideration. dQ is not, of course, a perfect differential. From this equation one obtains

$$\tfrac{1}{2}mn\frac{d}{dt}(u^2)+n\frac{dU}{dt}=n\frac{dQ}{dt}-\frac{\partial}{\partial\mathbf{x}}\cdot(\mathsf{p}\,.\,\mathbf{u}), \tag{3·17}$$

which may be regarded as the definition of the thermodynamic pressure tensor p in symbolic language. The corresponding equation for the energy of the macroscopic motion, obtained by multiplying (3·1) by $\mathbf{u}$, is

$$\tfrac{1}{2}mn\frac{d}{dt}(u^2)+\mathbf{u}\,.\left(\frac{\partial}{\partial\mathbf{x}}\cdot\mathsf{p}_1\right)=0 \tag{3·18}$$

since $\mathbf{F}$ vanishes. Subtracting from (3·17), one obtains

$$n\frac{dQ}{dt}=n\frac{dU}{dt}+\mathsf{p}:\left(\frac{\partial}{\partial\mathbf{x}}\mathbf{u}\right)+\mathbf{u}\,.\left\{\frac{\partial}{\partial\mathbf{x}}\cdot(\mathsf{p}-\mathsf{p}_1)\right\}, \tag{3·19}$$

or, by virtue of (3·14) and the equation of continuity,

$$n\frac{dQ}{dt}=n\frac{dU}{dt}+np\frac{d}{dt}\left(\frac{1}{n}\right)-\mathbf{u}\,.\frac{\partial\pi}{\partial\mathbf{x}} \tag{3·20}$$

on the assumption that the viscosity may be neglected; the conditions of validity of this hypothesis will be examined in § 5.

To complete the present account of non-uniform thermodynamics, it is necessary to add only the definition of the thermodynamic temperature T by means of the equation

$$\frac{dQ}{dt}=T\frac{dS}{dt}, \tag{3·21}$$

where S is the entropy per molecule, regarded as a space function of the three independent variables n, T and π. (3·21) is the customary formulation of the second law of thermodynamics, which, as was shown in part IV, goes over unchanged into the quantum theory: the necessity for this is, of course, obvious from the physical point of view. On the other hand, it is clear from (3·20) that the first law of thermodynamics does not hold in its usual sense for a quantum liquid; this does not imply any violation of the law of conservation of energy, which is actually implicit in its derivation, but is symptomatic of the existence of a curious interconnexion between the thermal energy and the energy of macroscopic motion.

4. THE ELEMENTARY THERMAL PROCESSES IN HE II

It is now expedient to consider the application of the general theory of the previous section to the three simple processes in fluids which are of special physical significance.

Consider first the quasi-static process in which a transition is made from one uniform state to another by the infinitely slow variation of an external parameter l. Such ideal conditions never occur naturally, but are of special importance inasmuch as they are the only kind to which ordinary reversible thermodynamics is strictly applicable. As the number density is a function of l only, the equation of continuity leads to

$$n'(l)\frac{dl}{dt}+n(l)\frac{\partial}{\partial \mathbf{x}}.\mathbf{u}=0, \quad u=u_0-x\frac{n'(l)}{n(l)}\frac{dl}{dt}, \tag{4·1}$$

assuming that the motion is in the x-direction. As $(dl/dt)^2$ and d^2l/dt^2 are completely negligible, the equation of motion (3·1) shows merely that p_1 and hence π are uniform. Thus (3·20) reduces to

$$dQ=dU+pd\left(\frac{1}{n}\right), \tag{4·2}$$

which is the ordinary law of thermodynamics.

When, however, steady processes, which are of frequent physical occurrence, are considered, one has $\mathbf{u}.\frac{\partial\pi}{\partial\mathbf{x}}=\frac{d\pi}{dt}$, so that (3·20) takes the form

$$dQ=dU+pd\left(\frac{1}{n}\right)-\frac{1}{n}d\pi \tag{4·3}$$

on multiplication by the time-differential. Part of any thermal energy communicated is absorbed by the macroscopic motion, as may be seen from (3·18), which leads to

$$d(\tfrac{1}{2}mu^2)+\frac{1}{n}dp=-\frac{1}{n}d\pi. \tag{4·4}$$

The equation (4·3) may be written in the form

$$TdS=d\left(U-\frac{\pi}{n}\right)+p_1d\left(\frac{1}{n}\right), \tag{4·5}$$

showing that the flowing helium behaves thermodynamically as if its internal energy were $U-\pi/n$, and the pressure were the kinetic pressure. The anomalous internal energy $-\pi/n$ gives rise to an anomalous specific heat additional to the normal specific heat, the additional energy supplied being converted into energy of motion. The motion may be visible, as in the transfer effect and in the flow through capillaries, or may take the form of closed currents analogous to the persistent currents in superconductors. It will be seen that such motion generally occurs only near the boundary walls containing the fluid, and takes another form inside the liquid. One may further rewrite (4·5) in the form

$$d\Phi=\frac{1}{n}dp_1-SdT, \tag{4·6}$$

where $\Phi=U+p/n-TS$ is the thermodynamic potential per molecule. For simultaneous mechanical and thermal equilibrium, one has $d\Phi=0$, since $dp_1=dT=0$, and in a purely thermomechanical process, $d\Phi=0$, so that

$$\frac{dp_1}{dT}=nS. \tag{4·7}$$

A third type of process, again of major physical interest, is that of wave propagation. There the number density may be taken to be a function of $t-v^{-1}\mathbf{k}.\mathbf{x}$ only, so that

$$v\frac{\partial n}{\partial \mathbf{x}}+\mathbf{k}\frac{\partial n}{\partial t} = 0, \tag{4·8}$$

where $\mathbf{k}$ is a unit vector in the direction of propagation, and v is the velocity of propagation. More complex types of wave disturbance may be obtained by the superposition of elementary waves of this type. From the substitution of (4·8) into the equation of continuity it follows that

$$\frac{\partial}{\partial \mathbf{x}}.(n\mathbf{u}-nv\mathbf{k}) = 0, \quad \mathbf{u} = v\left(1-\frac{n_0}{n}\right)\mathbf{k}, \tag{4·9}$$

where n_0 is a constant equal to the density at any place where the motion associated with the wave vanishes. Also

$$\left.\begin{aligned} n\frac{\partial \mathbf{u}}{\partial t} &= v\frac{n_0}{n}\frac{\partial n}{\partial t}\mathbf{k} = -v^2\frac{n_0}{n}\frac{\partial n}{\partial \mathbf{x}}, \\ n\mathbf{u}.\frac{\partial}{\partial \mathbf{x}}\mathbf{u} &= v^2\frac{n_0}{n}\left(1-\frac{n_0}{n}\right)\frac{\partial n}{\partial \mathbf{x}} = v^2\frac{n_0}{n}\frac{\partial n}{\partial \mathbf{x}}+v^2n_0^2\frac{\partial}{\partial \mathbf{x}}\left(\frac{1}{n}\right), \end{aligned}\right\} \tag{4·10}$$

so that the equation of motion leads to

$$\frac{\partial}{\partial \mathbf{x}}\left(mv^2\frac{n_0^2}{n}+p_1\right) = 0, \quad mv^2\frac{n_0^2}{n}+p_1 = c, \quad \text{a constant.} \tag{4·11}$$

Differentiating with respect to the density, one obtains

$$v^2 = \frac{1}{m}\frac{n^2}{n_0^2}\frac{dp_1}{dn}. \tag{4·12}$$

This is the ordinary formula for the velocity of sound if the propagation process is adiabatic, a condition which is obviously satisfied if the disturbance is purely mechanical, but not in thermal processes where the additional variable π comes into play. It is on the latter type of wave propagation that interest is now mainly centred.

For any thermodynamical variable, such as U, one has

$$\frac{\partial U}{\partial t} = -v\mathbf{k}.\frac{\partial U}{\partial \mathbf{x}} = \frac{n}{n_0-n}\mathbf{u}.\frac{\partial U}{\partial \mathbf{x}} = \frac{n}{n_0}\frac{dU}{dt}, \tag{4·13}$$

so that, for wave propagation, (3·20) leads to the law

$$dQ = dU+pd\left(\frac{1}{n}\right)-\frac{n_0-n}{n_0n}d\pi \tag{4·14}$$

or

$$TdS = d\left(U+\frac{\pi}{n_0}-\frac{\pi}{n}\right)+p_1d\left(\frac{1}{n}\right). \tag{4·15}$$

The anomalous specific heat is in this instance $\frac{d}{dT}\left(\frac{\pi}{n_0}-\frac{\pi}{n}\right)$, which us much smaller than that found for steady motion if the difference n_0-n is small. The effect of superposing a large number of waves with different directions, shapes and amplitudes, however, is to reduce the density considerably. This is evident from (4·9), on the understanding that the macroscopic velocity is in the opposite direction to the wave vector when the latter is identified with the direction of flow of heat. It thus appears that nearly all the anomalous specific heat in He II is accounted for by the presence of thermal waves, which are also responsible for the sudden decrease of the density below the λ-point.

To clarify the picture presented by the thermal waves, consider the effect of the transmission of a pulse of length l along a tube of cross-section σ, containing the liquid initially in its ground state, at rest. If, during the passage of the pulse, the value of π is reduced from ϖ to $\varpi - \Delta\pi$, it is clear from (4·14) that the heat transmitted is $(1-n/n_0)\,l\sigma\Delta\pi$, neglecting the small contribution $pd(1/n)$. From (4·9), on the other hand, one sees that the volume $uv^{-1}l\sigma = (n_0/n-1)\,l\sigma$ is transferred in the opposite direction. The whole process may occur isothermally and without variation of the pressure. This was found experimentally by Kapitza (1941*b*), who also measured the ratio $R = \Delta\pi/mn$ of the heat to the mass transfer. From (4·14), which for waves of small amplitude should be rewritten in the form

$$TdS = dU + (p+\Delta\pi)\,d\left(\frac{1}{n}\right), \tag{4·16}$$

one obtains further

$$R = \frac{T}{mn}\frac{dS}{d(1/n)}. \tag{4·17}$$

This formula is exact. The approximate formula $mR = TS$ which has been found empirically follows from the integral

$$TS = U + \frac{p+\Delta\pi}{n} + TS' \tag{4·18}$$

of (4·16) if one observes that nearly all the entropy is associated with the thermal wave, and hence $TS - \Delta\pi/n$ is very small.

Kapitza's measurements allow one to infer the variation with temperature of π for the normal state of the liquid, on the presumption that

$$\Delta\pi = nTS, \quad \varpi = nTS_0, \quad \pi = nT(S-S_0), \tag{4·19}$$

where S_0 is the maximum entropy associated with the thermal waves. The resulting curve is shown in figure 2, together with the linear curve for the ground state. From the point of view of London, Tisza, and Landau, an artificial separation of liquid helium can be made into a 'superfluid' and a 'normal' component; this could be expressed in terms of the present theory by writing $\pi = \varpi(n_s/n)$ and $\Delta\pi = \varpi(n_n/n)$, where n_s and n_n are the superfluid and normal densities respectively. This analogy between the theories suggests that the difference may be merely one of interpretation; physically absurd conclusions, however, result from the two-fluid conception, quite apart from its lack of theoretical justification, and it must therefore be discarded in favour of the more fundamental outlook.

The existence of the thermal waves in liquid helium in bulk was demonstrated experimentally by Peshkov (1944, 1946), who succeeded in evoking a train of waves with weak thermomechanical properties by passing an alternating current through an electrical heater immersed in the liquid, and measured the velocity of propagation by a resonance method. The conditions for the propagation of these waves, which may be inferred from (4·6) and (4·15), are

$$d\Phi = 0 \quad \text{and} \quad d\pi = 0; \tag{4·20}$$

it follows that

$$\frac{dp_1}{dT} = nS, \quad \frac{d}{dn}\{n(S-S_0)\} = 0, \tag{4·21}$$

if one neglects the small variation of temperature with density, and that the velocity of propagation of the waves is given by

$$mv^2 = \frac{dp_1}{dn} = S(S_0-S)\frac{dT}{dS}. \tag{4·22}$$

Tisza (1947) has shown that excellent numerical agreement with the experimentally observed velocity of 'second sound' can be obtained from this formula, using Kapitza's measurements of S. His method of derivation of this formula, on the basis of the two-fluid theory, is, however, complex, and open to objection from the present point of view.

5. VISCOSITY AND THERMAL CONDUCTION IN He II

It has been supposed in the foregoing that there are certain processes, other than quasi-static processes, in quantum liquids, for which the effect of viscosity can be neglected. From equation (3·2) it follows that viscosity is mainly due to the existence of the non-diagonal elements of the potential pressure l_1, defined by

$$\frac{\partial}{\partial \mathbf{x}^{(1)}}\cdot \mathsf{l}_1 = -\int \frac{\phi'(r)}{r}\,\mathbf{r}\,n_2(\mathbf{r},\mathbf{x})\,d\mathbf{x}^{(2)}, \tag{5·1}$$

where $\mathbf{r} = \mathbf{x}^{(2)} - \mathbf{x}^{(1)}$ and $\mathbf{x} = \frac{1}{2}(\mathbf{x}^{(1)} + \mathbf{x}^{(2)})$. The validity of the classical expansion in powers of the space gradients of density, temperature, and velocity is made rather dubious by the presence of waves and large velocities; fortunately, however, the equation (5·1) can be solved exactly by the operational methods of quantum mechanics. Since

$$e^{\frac{1}{2}\iota\mathbf{p}^{(1)}\cdot\mathbf{r}/\hbar}\,n_2(\mathbf{r},\mathbf{x}^{(1)}) = n_2(\mathbf{r},\mathbf{x}^{(1)}+\tfrac{1}{2}\mathbf{r}) = n_2(\mathbf{r},\mathbf{x}), \tag{5·2}$$

the equation reduces to

$$\frac{\iota}{\hbar}\mathbf{p}^{(1)}\,.\,\mathsf{l}_1 = -\int \frac{\phi'(r)}{r}\,\mathbf{r}\,e^{\frac{1}{2}\iota\mathbf{p}^{(1)}\cdot\mathbf{r}/\hbar}\,n_2(\mathbf{r},\mathbf{x}^{(1)})\,d\mathbf{r}. \tag{5·3}$$

Also, because $n_2(\mathbf{r},\mathbf{x}) = n_2(-\mathbf{r},\mathbf{x})$, the coefficient $(\iota/\hbar)\,\mathbf{p}^{(1)}$ can be cancelled from this equation, leaving

$$\mathsf{l}_1 = -\frac{1}{2}\int \frac{\phi'(r)}{r}\,\mathbf{r}\mathbf{r}\,\frac{\sin(\frac{1}{2}\mathbf{p}^{(1)}\,.\,\mathbf{r}/\hbar)}{\frac{1}{2}\mathbf{p}^{(1)}\,.\,\mathbf{r}/\hbar}\,n_2(\mathbf{r},\mathbf{x}^{(1)})\,d\mathbf{r}. \tag{5·4}$$

This leads to the classical approximation formulae if the ordinary expansion procedure is followed. One infers that the viscosity will vanish completely in the liquid bulk only if $n_2(\mathbf{r},\mathbf{x}^{(1)})$ is independent of $\mathbf{x}^{(1)}$; on the other hand, since n_2 is abnormally small in He II near the potential minimum, the viscosity will also be small, though finite. The experimental evidence available on the damping of mechanical motion in the liquid supports this conclusion. Also, as effectively uniform conditions are associated with the thermal waves investigated in the previous section, these waves will be quite undamped. The reversible processes in capillaries and narrow slits are accounted for entirely in terms of such waves, and viscosity need therefore be taken into account only to the extent to which these processes are often adulterated by the simultaneous operation of ordinary irreversible processes.

It is of interest to re-examine the derivation of Poiseuille's formula for the steady flow of a quantum liquid in a capillary. If r is the radial and x the axial co-ordinate, one has from (3·1) and (3·2)

$$\frac{dp_1}{dx}+\frac{1}{r}\frac{d}{dr}(rp_{xr}) = 0, \quad p_{xr} = -\mu\frac{du_x}{dr} = -\frac{r}{2}\frac{dp_1}{dx}, \tag{5·5}$$

where p_{xr} is the non-diagonal element of the pressure tensor, showing that the rate of decrease of velocity from the centre to the circumference is diminished in the ratio dp_1/dp, compared with the classical theory, and actually vanishes if p_1 is constant along the tube. In general, the slip between the walls and the fluid will be only partial, and the integral of (5·5) is

$$u_x = \frac{1}{4\mu}\left(-\frac{dp_1}{dx}\right)(a_1^2-r^2), \tag{5·6}$$

where the constant of integration a_1 becomes infinite only in the limit $dp_1/dx \to 0$, when u_x must tend to the constant value $-v(1-n_0/n)$. Indeed, in Kapitza's experiments (1941 *a*) on the jet phenomenon produced by heating the liquid in a capillary sealed at one end, it is necessary to suppose that a_1 is less than the radius a of the capillary, as may be seen from the expression $\frac{\pi a^4}{8\mu}\left(-\frac{dp_1}{dx}\right)\left(\frac{2a_1^2}{a^2}-1\right)$ for the volume flow across the cross-section, which must almost vanish for a sealed capillary. These experiments were proved to be strongly irreversible, consistent with the large viscous forces required on theoretical grounds.

The laws of viscosity and thermal conduction, expressed in the form of (3·2) and (3·4) respectively, are not very useful because they contain reference to the 'kinetic' temperature and pressure alone. To formulate approximate laws with reference to the corresponding thermodynamic quantities, one is guided by the requirement that $\mathbf{q}$ and $\mathfrak{p}'$ should be linear in the space derivatives of T and $\mathbf{u}$, and must accordingly have the form

$$\mathbf{q} = -\lambda'\frac{\partial T}{\partial \mathbf{x}}+2\kappa\mathbf{u} : \overgroup{\left(\frac{\partial}{\partial \mathbf{x}}\mathbf{u}\right)}, \quad \mathfrak{p}' = -2\mu'\overgroup{\frac{\partial}{\partial \mathbf{x}}\mathbf{u}}+\frac{2\theta}{u^2}\overgroup{\frac{\partial T}{\partial \mathbf{x}}\mathbf{u}}. \tag{5·7}$$

The second term in each of these equations does not appear in the classical theory, for the reason that there, $\mathbf{u}$ may not occur explicitly, but they are not excluded and are indeed required by the quantum theory, where thermal and dynamical quantities are inseparably linked. The quantity κ may be called 'the coefficient of thermal mobility', and θ 'the modulus of thermal stress'.

For the complete discussion of irreversible phenomena, it is necessary to have the general solution of the equation (3·1) in steady conditions, when, together with the equation of continuity, it reduces to

$$mn\mathbf{u}\,.\frac{\partial}{\partial \mathbf{x}}\mathbf{u}+\frac{\partial}{\partial \mathbf{x}}\,.\,\mathbf{p}_1, \quad \frac{\partial}{\partial \mathbf{x}}\,.\,(n\mathbf{u}) = 0. \tag{5·8}$$

This pair of equations is integrable, with the solution

$$mn\mathbf{u}\mathbf{u}+\mathbf{p}_1 = \frac{\partial}{\partial \mathbf{x}}\times \mathbf{s}, \tag{5·9}$$

where $\mathbf{s}$ is an arbitrary tensor, fixed by the boundary conditons, or

$$mn\mathbf{u}\mathbf{u}+\mathbf{p}_1^0 = \frac{\partial}{\partial \mathbf{x}}\times \mathbf{s}^0, \quad \mathbf{p}_1' = \frac{\partial}{\partial \mathbf{x}}\times \mathbf{s}'. \tag{5·10}$$

The first of these equations can be satisfied by substituting for $\mathbf{s}^0$ an antisymmetrical tensor defined by $(s^0_{yz}, s^0_{zx}, s^0_{xy}) = \mathbf{t}\,.\,\mathbf{x}$ $(\mathbf{t}_{xy} = t_{yx}$, etc.) if the left-hand side is uniform, or otherwise by a more complicated integral expression. Then one has

$$mn\mathbf{u}\mathbf{u}+\mathbf{p}_1^0 = tr(\mathbf{t})\,\mathbf{1}-\mathbf{t}. \tag{5·11}$$

Now suppose the flow is in a slit bounded by two planes separated by a very small distance d and normal to the z-axis. Then k_{1zz}, which represents the mean square momentum in the z-direction, must, according to Heisenberg's uncertainty principle, be of order $(\hbar d^{-1})^2$. The relevance of this conclusion has been pointed out by Daunt & Mendelssohn (1946). Writing

$$p^0_{1zz} = p_1+\gamma(\hbar d^{-1})^2, \tag{5·12}$$

so that p_1 is the hydrostatic pressure in the infinite liquid, from (5·11) it follows, since $u_z = 0$, that

$$p_1+\gamma(\hbar d^{-1})^2 = tr(\mathbf{t})-t_{zz}. \tag{5·13}$$

This shows that $\mathbf{t}$ has a part of order d^{-2}, so that

$$\mathbf{t} = \mathbf{t}^0+(\hbar d^{-1})^2\,\mathbf{t}', \quad \gamma = t'_{xx}+t'_{yy}. \tag{5·14}$$

It is then clear from the equations

$$mnu_x^2+p_1 = tr(\mathbf{t})-t_{xx}, \quad mnu_y^2+p_1 = tr(\mathbf{t})-t_{yy}, \tag{5·15}$$

that when d is very small, u_x and u_y are of order d^{-1}, being given, sufficiently nearly, by

$$(mn)^{\frac{1}{2}}\,u_x d = \hbar(t'_{yy}+t'_{zz})^{\frac{1}{2}}, \quad (mn)^{\frac{1}{2}}\,u_y d = \hbar(t'_{xx}+t'_{zz})^{\frac{1}{2}}. \tag{5·16}$$

The tensor $\mathbf{t}$ is determined by the boundary conditions, effectively the conditions beyond the end of the slit. This confirms the conclusion reached previously, that the volume flow through capillaries and narrow slits includes a component depending not on the width of the channel, but only on the thermal transfer between the two ends. There may, however, exist an independent component corresponding to the behaviour of normal liquids.

The same conclusions apply to the behaviour of the thin films associated with the transfer phenomenon, where the distance d is determined naturally instead of by physical constraints. The method of initial formation of these films is rather complex, involving thermal fluctuations within the liquid and perhaps also evaporation and condensation from the vapour. Once formed, however, the mechanism responsible for their perpetuation is precisely the same as that involved in the superfluid motion through capillaries and narrow slits.

The author wishes to acknowledge with sincerest thanks the important part played by the advice and encouragement which he has received from Professor Max Born in the development of the theory. He is indebted to Imperial Chemical Industries for the tenure of a Research Fellowship.

REFERENCES

de Boer, J. & Michels, A. 1939 *Physica, Haarlem*, **6**, 97.
Born, M. & Green, H. S. 1946 *Proc. Roy. Soc.* A, **188**, 10.
Born, M. & Green, H. S. 1947*a* *Proc. Roy. Soc.* A, **190**, 455.
Born, M. & Green H. S. 1947*b* *Proc. Roy. Soc.* A, **191**, 168.
Daunt, J. G. & Mendelssohn, K. 1939 *Proc. Roy. Soc.* A, **170**, 423.
Daunt, J. G. & Mendelssohn, K. 1946 *Phys. Rev.* **69**, 126.
Green, H. S. 1947 *Proc. Roy. Soc.* A, **189**, 103.
Kapitza, P. L. 1941*a* *J. Phys. Acad. Sci. U.S.S.R.* **4**, 181.
Kapitza, P. L. 1941*b* *J. Phys. Acad. Sci. U.S.S.R.* **5**, 59.
Landau, L. 1941 *J. Phys. Acad. Sci. U.S.S.R.* **5**, 71.
Landau, L. 1944 *J. Phys. Acad. Sci. U.S.S.R.* **8**, 1.
London, F. 1938 *Phys. Rev.* **54**, 947.
London, F. 1939 *J. Phys. Chem.* **43**, 49.
Peshkov, V. 1944 *J. Phys. Acad. Sci. U.S.S.R.* **8**, 381.
Peshkov, V. 1946 *J. Phys. Acad. Sci. U.S.S.R.* **10**, 389.
Tisza, L. 1938 *C.R. Acad. Sci., Paris*, **207**, 1035, 1186.
Tisza, L. 1940 *J. Phys. Radium* (ser. 8), **1**, 165, 350.
Tisza, L. 1947 *Phys. Rev.* **72**, 838.

Electroviscosity

I. The flow of liquids between surfaces in close proximity

By G. A. H. Elton, *Chemistry Department, Battersea Polytechnic*

(*Communicated by J. Kenyon, F.R.S.—Received* 20 *December* 1947)

The conditions governing the rate of flow of ionic and non-ionic liquids in narrow channels are examined, and it is demonstrated that in certain circumstances the electrical resistance to shear in ionic liquids, due to the effect of an electrical potential gradient on the ions of the surface layers, may assume considerable importance relative to the mechanical (viscous) resistance to shear. The results of this analysis are of important application in various fields, and offer an explanation of anomalous results obtained by numerous experimental workers.

1. Introduction

In recent years a number of workers have devoted considerable attention to the behaviour of thin films of liquid between two interfaces (liquid/solid, liquid/liquid or liquid/gas) when subject to forces tending to reduce the film thickness. For example, Derjaguin & Kussakov (1939*a*, *b*) and Derjaguin (1940) studied the behaviour of gas bubbles in close proximity to various solid surfaces in water, in dilute aqueous salt solutions, and in various non-aqueous liquids. Eversole & Lahr (1941*b*) studied the separation produced between quartz surfaces by the condensation of a water film between them, and Buzagh (1937) studied, *inter alia*, the sedimentation of glass disks in aqueous salt solutions. Generally, the conclusions reached were that such films, up to a thickness of several hundred Angstroms, exhibited certain anomalous properties. Derjaguin & Kussakov reported that if a pressure were applied to such a film, its thickness did not decrease indefinitely, but took up a definite equilibrium value corresponding to that pressure. In aqueous solutions, for any given pressure and surfaces, the equilibrium film thickness was a function of the ion concentration, decreasing as the ion concentration increased. For a water film between a water/hydrogen and a water/mica interface, the equilibrium thickness varied between 1700 A at 1000 dynes/cm.^{2} pressure to 300 A at 14,000 dynes/cm.^{2}.

The equilibrium distances of separation between two charged plane surfaces immersed in a liquid (and subject to no external forces which would tend to reduce the separation) have been calculated theoretically by Verwey & Overbeek (1946) and Levine & Dube (1939), and found to be of a much lower order (0 to 50 A); Schofield (1946) has performed calculations which lead to equilibrium distances of separation similar in some cases to those of Derjaguin & Kussakov, but fail to account for the effect of ion concentration.

Some of the experimental work in this field has been criticized by the writer (Elton 1946), and this criticism will be extended in a future publication. The present paper will be confined to the discussion of a phenomenon which is of crucial importance in systems of this type, and which has not previously been recognized by workers in this field. Preliminary experimental evidence for the existence of this effect is given in the paper immediately following.

2. VISCOSITY EFFECTS IN THIN LIQUID FILMS

When considering the stability of a liquid film subjected to pressure between two interfaces, it is essential first of all to examine the effect of ordinary viscous resistance to shear which will delay the attainment of any equilibrium state. The attainment of such a state may involve the lapse of a considerable period of time, depending on the size and shape of the interfaces and the initial and final distances of separation (see § 3). In addition, if the liquid is ionic and has a finite electrokinetic (ζ) potential at one or both of the interfaces, an additional retarding force must be taken into account.

Whenever an electrical double layer at an interface in an ionic liquid is sheared, a potential difference, known as a streaming potential, is set up in the plane of shear. This potential difference will tend to resist the flow of liquid owing to the electrical retarding force acting on the ions of the double layer. In § 4 it is shown that in thin films of liquid, this effect can assume great importance, with the result that the total resistance to approach of two surfaces in an ionic liquid may greatly exceed the corresponding resistance in a non-ionic liquid of the same bulk viscosity.

3. VISCOUS RESISTANCE TO APPROACH OF TWO SURFACES IN A NON-IONIC LIQUID

In 1886, Reynolds showed theoretically that the approach of two plane parallel disks of radius a, under a force mg in a liquid of bulk viscosity η was governed by the equation

$$t=\frac{3\pi\eta a^4}{4mg}\left(\frac{1}{h_2^2}-\frac{1}{h_1^2}\right), \tag{1}$$

where h_2 is the distance of separation between the two surfaces at time t, and h_1 is the initial separation.

Hardy & Bircumshaw (1925) give an expression (derived by G. I. Taylor) for the approach of a sphere of radius of curvature R to a large flat plate, under the action of a force mg in a liquid of bulk viscosity η:

$$t=\frac{6\pi\eta R^2}{mg}\ln\frac{h_1}{h_2}. \tag{2}$$

Both of these equations lead to infinite values of t for zero final separation, but equation (2) leads to values of a lower order of infinity than equation (1). That is to say, under equal forces, a sphere will approach a plate very much more rapidly than will a plate of the same order of size, immersed in the same liquid.

Differentiation of equation (1) leads to the expression

$$\frac{dh_2}{dt}=-\frac{2mgh_2^3}{3\pi\eta a^4}, \tag{3}$$

and differentiation of equation (2) leads to

$$\frac{dh_2}{dt}=-\frac{mgh_2}{6\pi\eta R^2}, \tag{4}$$

so that for parallel plates the velocity of approach is proportional to the cube of the distance of separation, while for a sphere approaching a plate, the velocity of approach is proportional to the first power of the distance of separation.

Before proceeding any further it must be emphasized that equations (1) to (4) must be applied only to cases involving a statistical number of molecules. For example, it would not be legitimate to apply these equations down to values of h_2 of say 10 A. In the absence of any complicating phenomena, however, it should be permissible to apply these equations down to quite small distances of separation (at least to 1000 A, possibly approximately to 100 A). If this is done, it is found that the rate of approach of parallel plates is very slow compared with the rate of approach of a sphere of similar dimensions to a plate. Table 1 gives the calculated times of approach of (*a*) a circular plate of radius 1 cm. to a similar plate, and (*b*) a sphere of radius 1 cm. to a large flat plate, in both cases under a force of 1000 dynes in a liquid of bulk viscosity 0·01 poise, and from an initial height of 1 cm. It is seen from table 1 that the time taken for a plate to reach a distance of separation of the order of a few hundred Angstroms is considerable. A distance of separation of 1000 A is reached in about 70 hr., and a distance of 100 A in about 7000 hr. The calculated rate of approach at the former distance is 2×10^{-11} cm./sec. (Such a value has, of course, no absolute significance in view of the above-mentioned limits of applicability of the equations.) On the other hand, the sphere has attained a very small distance in a very short time, and the rate of approach is still relatively rapid.

TABLE 1

distance of separation (cm.)	time for sphere (sec.)	time for plate (sec.)
1	0	0
1×10^{-1}	42×10^{-5}	$2{\cdot}4 \times 10^{-3}$
1×10^{-2}	84×10^{-5}	$2{\cdot}4 \times 10^{-1}$
1×10^{-3}	126×10^{-5}	$2{\cdot}4 \times 10^{1}$
1×10^{-4}	168×10^{-5}	$2{\cdot}4 \times 10^{3}$
1×10^{-5}	210×10^{-5}	$2{\cdot}4 \times 10^{5}$
1×10^{-6}	252×10^{-5}	$2{\cdot}4 \times 10^{7}$

4. THE ELECTROVISCOUS EFFECT

It is proposed in this section to derive equations expressing the apparent increase in viscosity of an ionic liquid (due to the resistance to shear of the double layer),

(*a*) in a cylindrical capillary tube,

(*b*) between two infinite parallel plates, and

(*c*) to use these equations in the derivation of the order of the electroviscous effect between two approaching parallel plates.

(*a*) *The electroviscous effect in a narrow cylindrical capillary*

In the derivation of this equation it is necessary to know the potential at any point inside a narrow capillary containing an ionic liquid, as a function of the distance of that point from the axis. It is proposed first of all to derive this relationship.*

(i) *The distribution of potential in a narrow cylindrical capillary containing an ionic liquid*

Consider a capillary tube of radius a, containing a solution of a uni-univalent electrolyte, of bulk ionic concentration n ions/unit volume. Then if the wall of the tube is at a potential ψ_0, and the potential is ψ at a point distance r from the axis, the Poisson equation relates the potential gradient at this point with the excess charge density ρ_r

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{d\psi}{dr}\right) = -\frac{4\pi}{\epsilon}\rho_r, \tag{5}$$

where ϵ is the dielectric constant, assumed constant throughout the liquid.

From the Boltzmann equation

$$\rho_r = -2ne\sinh\frac{e\psi}{\boldsymbol{k}T}, \tag{6}$$

where e = electronic charge, $\boldsymbol{k}$ = Boltzmann constant and T = absolute temperature. If $e\psi/\boldsymbol{k}T$ is small, we may write $\sinh(e\psi/\boldsymbol{k}T) = e\psi/\boldsymbol{k}T$. (This assumption limits the application of this treatment to small values of ψ, since $e\psi/\boldsymbol{k}T = 1$ when $\psi \simeq 25$ mV.) Combining equations (5) and (6), and approximating,

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{d\psi}{dr}\right) = \frac{8\pi ne^2\psi}{\epsilon\boldsymbol{k}T}. \tag{7}$$

Let

$$\left(\frac{\epsilon\boldsymbol{k}T}{8\pi ne^2}\right)^{\frac{1}{2}} = \lambda,$$

where λ is the Debye double layer thickness. Therefore

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{d\psi}{dr}\right) = \frac{\psi}{\lambda^2}. \tag{8}$$

This equation may be solved as a series:

$$\psi = Ar^2\left(1 + \frac{r^2}{16\lambda^2} + \frac{r^4}{36\lambda^2} + \dots\right), \tag{9}$$

where A is a constant. Since $\psi = \psi_0$ when $r = a$, we may write, when r is small compared with r/λ,

$$\psi = \frac{\psi_0 r^2}{a^2}\left(\frac{16\lambda^2 + r^2}{16\lambda^2 + a^2}\right). \tag{10}$$

(ii) *Derivation of the electroviscous equation*

Consider a cylindrical capillary tube of radius a, having a pressure gradient P applied along its axis, forcing through it an ionic liquid of bulk viscosity η. If a finite ζ-potential exists at the liquid/solid interface, a streaming potential gradient E is set up.

Considering an annulus of liquid of radius r and thickness δr situated symmetrically with respect to the axis of the tube, and having unit length, then

$$\text{Viscous retarding force on this annulus} = 2\pi\eta\frac{d}{dr}\left(r\frac{du}{dr}\right)\delta r, \tag{11}$$

where u is the velocity of flow of the liquid, and

$$\text{Force on annulus due to the pressure difference} = 2\pi r\,\delta r\,P. \tag{12}$$

In the absence of any other forces at equilibrium, the normal Poiseuille relationship is obtained after equating these two terms. However, since the liquid is an ionic one, the ions of the diffuse double layer will suffer a retarding force due to the potential gradient against which they are being forced. Thus

$$\text{Electrical retarding force: } E\,.\,2\pi r\,\delta r\,\rho_r. \tag{13}$$

At equilibrium, i.e. when the liquid is moving without acceleration through the tube,

$$-2\pi\eta\frac{d}{dr}\left(r\frac{du}{dr}\right)\delta r - E\,.\,2\pi r\,\delta r\,\rho_r = 2\pi r\,\delta r\,P. \tag{14}$$

From Poisson's equation $$\frac{4\pi\rho_r}{\epsilon} = -\frac{1}{r}\frac{d}{dr}\left(r\frac{d\psi}{dr}\right). \tag{15}$$

Hence $$-\eta\frac{d}{dr}\left(r\frac{du}{dr}\right) + \frac{\epsilon E}{4\pi}\frac{d}{dr}\left(r\frac{d\psi}{dr}\right) = Pr. \tag{16}$$

Integrating $$-\eta r\frac{du}{dr} + \frac{\epsilon E}{4\pi}r\frac{d\psi}{dr} = \frac{Pr^2}{2} + A, \tag{17}$$

where A is the integration constant. Integrating again

$$-\eta u + \frac{\epsilon E\psi}{4\pi} = \frac{Pr^2}{4} + A\ln r + B, \tag{18}$$

where B is the second integration constant. Since at the axis of the tube ($r = 0$), u is not equal to minus infinity, $A = 0$. If at the walls of the tube, $u = 0$, and $\psi = \psi_0$,

$$B = \frac{\epsilon E\psi_0}{4\pi} - \frac{Pa^2}{4}. \tag{19}$$

Hence $$-\eta u = \frac{P(r^2 - a^2)}{4} + \frac{\epsilon E}{4\pi}(\psi_0 - \psi). \tag{20}$$

The additional term due to the electrical retarding force will only become important in extremely narrow capillaries. The effect will be to increase the apparent viscosity of the liquid. If η_a is the apparent viscosity in the tube such that

$$u\eta_a = \frac{P(a^2 - r^2)}{4}, \tag{21}$$

we can deduce an expression for η_a as follows:

If Q is the volume of liquid flowing through the tube in unit time, then

$$Q = \int_0^a 2\pi r u\,dr = \frac{\pi}{2\eta}\int_0^a\left[Pr(a^2 - r^2) - \frac{\epsilon Er}{4\pi}(\psi_0 - \psi)\right]dr. \tag{22}$$

As already shown, for a narrow tube we may write

$$\psi = \frac{\psi_0 r^2}{a^2}\left(\frac{16\lambda^2 + r^2}{16\lambda^2 + a^2}\right). \tag{10}$$

Thus for unit length of tube

$$Q = \frac{\pi}{2\eta}\int\left[Pr(a^2-r^2)-\frac{\epsilon E\psi_0 r}{\pi}\left\{1-\frac{r^2}{a^2}\left(\frac{16\lambda^2+r^2}{16\lambda^2+a^2}\right)\right\}\right]dr \tag{23}$$

$$= \frac{\pi}{2\eta}\left[\frac{Pa^2r^2}{2}-\frac{Pr^4}{4}-\frac{\epsilon E\psi_0 r^2}{2\pi}\left\{1-\frac{r^2}{3a^2}\left(\frac{24\lambda^2+r^2}{16\lambda^2+a^2}\right)\right\}\right]_0^a \tag{24}$$

$$= \frac{\pi}{2\eta}\left[\frac{Pa^4}{4}-\frac{\epsilon E\psi_0 a^2}{2\pi}\left\{1-\frac{(24\lambda^2+a^2)}{3(16\lambda^2+a^2)}\right\}\right]. \tag{25}$$

If the reduced rate of flow results in an apparent viscosity η_a, where

$$Q = \frac{\pi Pa^4}{8\eta_a} \quad \text{and} \quad E = +\frac{\zeta\epsilon P}{4\pi\eta_a\kappa},$$

κ being the average specific conductivity of the liquid in the tube, then putting $\psi = \zeta$, where ζ is the electrokinetic potential of the solution against the walls of the tube,

$$\frac{\pi Pa^4}{8\eta_a} = \frac{\pi Pa^4}{8\eta}-\frac{\zeta^2\epsilon^2Pa^2}{16\pi\eta\eta_a\kappa}\left\{1-\frac{(24\lambda^2+a^2)}{3(16\lambda^2+a^2)}\right\}. \tag{26}$$

Therefore

$$\eta_a = \eta+\frac{\zeta^2\epsilon^2}{2\pi^2a^2\kappa}\left\{1-\frac{(24\lambda^2+a^2)}{3(16\lambda^2+a^2)}\right\}. \tag{27}$$

If $a \gg \lambda$ this equation reduces to

$$\eta_a = \eta+\frac{\zeta^2\epsilon^2}{3\pi^2a^2\kappa}. \tag{28}$$

When (1) a is large, and/or (2) ζ is zero or very small, and/or (3) κ is large, then equation (28) reduces to

$$\eta_a = \eta. \tag{29}$$

(iii) *Approximate derivation of the electroviscous equation for a narrow cylindrical capillary*

We may regard the apparent increase in viscosity in narrow tubes as being due to a 'back electroendosmosis' effect produced by the streaming potential gradient set up. This produces a reduction in the rate of flow, and an increase in the apparent viscosity.

An approximate equation may be derived on this basis from the usual expressions for streaming potential and electroendosmosis in cylindrical capillaries as follows:

The streaming potential in a wide tube is given by

$$E = +\frac{P\epsilon\zeta}{4\pi\eta\kappa}. \tag{30}$$

For electroendosmosis in a wide tube

$$E' = +\frac{\pi a^2P'}{2\epsilon\zeta}, \tag{31}$$

where E' is the applied potential gradient and P' is the electroendosmosis pressure set up.

In a narrow tube, where an apparent viscosity η_a is obtained due to electroviscous resistance, from equation (28) we have

$$\eta_a = +\frac{P\epsilon\zeta}{4\pi E\kappa}. \tag{32}$$

Since the relative difference between the apparent and bulk viscosities is due to the 'back-electroendosmosis' effect of the streaming potential gradient,

$$\frac{\eta_a-\eta}{\eta_a}=+\frac{2\zeta E'\epsilon}{\pi a^2 P'}. \tag{33}$$

Combining equations (32) and (33), and putting $E = E'$, and $P = P'$, we arrive at

$$\eta_a=\eta+\frac{\zeta^2\epsilon^2}{2\pi^2a^2\kappa}, \tag{34}$$

which is identical in form with equation (28), but differs in the value of the constant. This difference is due to the simplifying assumption made in the derivation of equations (30) and (31), namely, that a plane parallel double layer of negligible thickness exists at the walls of the tube. Subject to the limitations imposed by these assumptions, this equation (34) may be applied for any value of ζ.

(*b*) *The electroviscous effect between two infinite parallel plates*

Again we require to know how the potential at any point between the two plates varies with the distance of that point from one of them.

(i) *The distribution of potential between two parallel plates*

Consider two parallel plates separated by a distance $2h$, containing a solution of a uni-univalent electrolyte of ionic concentration n. Then from the Poisson equation

$$\frac{d^2\psi}{dx^2}=-\frac{4\pi\rho_x}{\epsilon}, \tag{35}$$

where ψ is the potential at a point at distance x from one of the plates, ρ_x is the excess charge density at the point, and ϵ is the dielectric constant, assumed constant throughout the liquid.

From the Boltzmann equation

$$\rho_x=-2ne\sinh\frac{e\psi}{kT}. \tag{36}$$

If $e\psi/kT$ is small, we may write $e\psi/kT=\sinh e\psi/kT$ (vide supra). Combining equations (35) and (36), and approximating

$$\frac{d^2\psi}{dx^2}=\frac{8\pi ne^2}{\epsilon kT}\psi. \tag{37}$$

Let $$\left(\frac{\epsilon kT}{8\pi ne^2}\right)^{\frac{1}{2}}=\lambda.$$

Then $$\frac{d^2\psi}{dx^2}=\frac{\psi}{\lambda^2}. \tag{38}$$

Integrating $$\frac{d\psi}{dx}=\left(\frac{\psi^2}{\lambda^2}+A\right)^{\frac{1}{2}}, \tag{39}$$

if $\psi=d\psi/dx=0$ when $x=h$, then $A=0$. Therefore

$$\frac{d\psi}{dx}=\pm\frac{\psi}{\lambda}. \tag{40}$$

$d\psi/dx$ is negative when $x < h$, and positive when $h < x < 2h$. Integrating

$$\frac{x}{\lambda} = \pm \ln \psi + B. \tag{41}$$

When $x = 0$ or $2h$, $\psi = \psi_0$. Therefore

$$B = \ln \psi_0 \quad \text{or} \quad \frac{2h}{\lambda} - \ln \psi_0.$$

Therefore
$$\frac{x}{\lambda} = \ln \frac{\psi_0}{\psi} \quad \text{or} \quad \frac{x-2h}{\lambda} = \ln \frac{\psi}{\psi_0}. \tag{42}$$

Therefore
$$\psi = \psi_0 e^{-x/\lambda} \quad \text{or} \quad \psi_0 e^{(x-2h)/\lambda}. \tag{43}$$

(ii) *Derivation of the electroviscous equation for flow between parallel plates*

Let two similar parallel plates of infinite extent be separated by a distance $2h$, let an ionic liquid of normal viscosity η be forced between them under a pressure gradient P, and let a streaming potential gradient E be set up in the plane of flow.

Consider a lamina of liquid of unit length and breadth, and of thickness δx, parallel to the plates and at a distance x from one of them. Then

$$\text{Viscous retarding force on this lamina} = \eta\, \delta x \frac{d^2u}{dx^2}, \tag{44}$$

where u is the velocity of flow of the liquid,

$$\text{Force on the lamina due to the pressure gradient} = P\delta x. \tag{45}$$

An additional force, namely, the electroviscous force on the ions of the double layer, must now be taken into account. If the

$$\text{Electrical retarding force on the lamina} = E\rho_x \delta x, \tag{46}$$

then, at equilibrium,
$$-\eta\, \delta x \frac{d^2u}{dx^2} - E\rho_x \delta x = P\delta x.. \tag{47}$$

By Poisson's equation
$$\frac{d_2\psi}{dx^2} = \frac{4\pi\rho_x}{\epsilon}. \tag{48}$$

Hence
$$-\eta \frac{d}{dx}\left(\frac{du}{dx}\right) + \frac{\epsilon E}{4\pi} \frac{d}{dx}\left(\frac{d\psi}{dx}\right) = P. \tag{49}$$

Integrating
$$-\eta \frac{du}{dx} + \frac{\epsilon E}{4\pi} \frac{d\psi}{dx} = Px + A, \tag{50}$$

where A is the integration constant. Integrating again

$$-\eta u + \frac{\epsilon E \psi}{4\pi} = \frac{Px^2}{2} + Ax + B, \tag{51}$$

where B is the second integration constant. Now at the plates $u = 0$ and $\psi = \psi_0$, $x = 0$ or $2h$. Therefore

$$B = \frac{\epsilon E \psi_0}{4\pi} \quad \text{and} \quad A = -Ph.$$

Therefore
$$\eta u = -\frac{Px^2}{2} + \frac{\epsilon E}{4\pi}(\psi - \psi_0) + Phx, \tag{52}$$

$$u = \frac{1}{\eta}\left[\frac{Px}{2}(2h-x) + \frac{\epsilon E}{4\pi}(\psi - \psi_0)\right]. \tag{53}$$

The volume of liquid flowing per second is

$$Q = \int_0^{2h} u\,dx = \frac{1}{\eta}\int_0^{2h}\left[\frac{Px}{2}(2h-x) + \frac{\epsilon E}{4\pi}(\psi - \psi_0)\right]dx. \tag{54}$$

Substituting for ψ from equation (43), we have

$$Q = \frac{1}{\eta}\int_0^{h}\left[\frac{Px}{2}(2h-x) - \frac{\epsilon E\psi_0}{4\pi}(1 - e^{-x/\lambda})\right]dx + \frac{1}{\eta}\int_h^{2h}\left[\frac{Px}{2}(2h-x) - \frac{\epsilon E\psi_0}{4\pi}(1 - e^{(x-2h)/\lambda})\right]dx \tag{55}$$

$$= \frac{1}{\eta}\left[\frac{Px^2h}{2} - \frac{Px^3}{6}\right]_0^{2h} - \frac{\epsilon E\psi_0}{4\pi\eta}\left\{\left[x + \lambda e^{-x/\lambda}\right]_0^h + \left[x - \lambda e^{(x-2h)/\lambda}\right]_h^{2h}\right\} \tag{56}$$

$$= \frac{1}{\eta}\left[\frac{2Ph^3}{3} - \frac{\epsilon E\psi_0}{2\pi}(h - \lambda + \lambda e^{-h/\lambda})\right]. \tag{57}$$

Let the reduced rate of flow result in an apparent viscosity η_a, where

$$Q = \frac{2Ph^3}{3\eta_a} \quad \text{and} \quad E = +\frac{\zeta P\epsilon}{8\pi\eta_a\kappa}.$$

Putting $\zeta = \psi_0$
$$\frac{2Ph^3}{3\eta_a} = \frac{2Ph^3}{3\eta} - \frac{P\epsilon^2\zeta^2}{16\pi^2\kappa\eta\eta_a}(h - \lambda + \lambda e^{-h/\lambda}), \tag{58}$$

$$\eta_a = \eta + \frac{3\epsilon^2\zeta^2}{32\pi^2\kappa h^3}(h - \lambda + \lambda e^{-h/\lambda}). \tag{59}$$

When $h \gg \lambda$, this reduces to
$$\eta_a = \eta + \frac{3\epsilon^2\zeta^2}{32\pi^2\kappa h^2}. \tag{60}$$

(iii) *Approximate derivation of the electroviscous equation for flow between parallel plates*

Equation (60) may be derived by combining the approximate equations for electroendosmosis and streaming potential between parallel plates, as follows:

If η is the bulk viscosity of the liquid, and η_a the apparent viscosity, taking into account the electroviscous effect, P is the applied pressure gradient, and E the resulting streaming potential gradient, then we may show by the usual methods that

$$\eta_a = +\frac{P\zeta\epsilon}{8\pi E\kappa}, \tag{61}$$

and for electroendosmosis
$$\frac{\eta_a - \eta}{\eta_a} = +\frac{3\epsilon\zeta E'}{4\pi h^2 P'}, \tag{62}$$

where E' is an applied potential gradient, P' the resultant electroendosmotic pressure gradient, and $2h$ the distance of separation of the plates.

Combining equations (61) and (62), putting $E = E'$, and $P = P'$, we obtain

$$\eta_a = \eta + \frac{3\epsilon^2\zeta^2}{32\pi^2\kappa h^2}. \tag{63}$$

This equation is identical with equation (60), subject to the limitations imposed in the derivation of equations (61) and (62), namely, that the double-layer thickness is very small compared with the distance of separation of the plates, this equation (63) may be applied for any value of ζ.

(c) *The approach of two parallel plates in an ionic liquid*

By analogy with the two previous examples, we may write for the approach of two parallel plates in an ionic liquid

$$\frac{dh_2}{dt} = -\frac{2mgh_2^3}{3\pi\eta_a a^4}, \tag{64}$$

where $\eta_a = \eta + A\epsilon^2\zeta^2/h^2\kappa$ approx., A being a constant, and h_2 the distance of separation of the plates at time t. Integrating equation (64) from h_1 to h_2

$$t = \frac{3\pi\eta a^4}{4mg}\left(\frac{1}{h_2^2} - \frac{1}{h_1^2}\right) + \frac{C\epsilon^2\zeta^2}{k}\left(\frac{1}{h_2^4} - \frac{1}{h_1^4}\right), \tag{65}$$

where C is a constant, equal to $3\pi Aa^4/8mg$. The final term in equation (65) does not become important until h_2 is very small (less than 10,000 A usually). We may obtain some idea of the order of the electroviscous term if we put $A = \frac{3}{8}\pi^2$, as in equation (60) for the flow of an ionic liquid between two infinite fixed parallel planes separated by a distance $2h$. C in equation (65) then becomes $9a^4/64\pi mg$. Calculations made on this basis lead to the results shown in table 2, which gives the calculated times of approach under a force of 1000 dynes of a disk of radius 1 cm. to a similar disk (*a*) in a non-ionic liquid of bulk viscosity 0·01 poise, and (*b*) in an ionic liquid ($\zeta = 100\,\text{mV}$, $\epsilon = 80$ and $k = 10^{-6}\,\text{ohm}^{-1}\text{cm.}^{-1}$) of bulk viscosity 0·01 poise. At a separation of 1×10^{-5} cm., the calculated rate of approach in the non-ionic liquid is 2×10^{-11} cm./sec. and in the ionic liquid is 6×10^{-13} cm./sec. (subject to the limits of applicability already discussed). The time taken to fall from 1×10^{-5} to 9×10^{-6} cm. in the latter case is $1{\cdot}6 \times 10^6$ sec. (about 450 hr.), whereas in the non-ionic liquid the time is 5×10^4 sec. (about 15 hr.).

TABLE 2

distance of separation (cm.)	time in ionic liquid (sec.)	time in non-ionic liquid (sec.)
10^{-4}	$2{\cdot}5 \times 10^3$	$2{\cdot}4 \times 10^3$
5×10^{-5}	$1{\cdot}1 \times 10^4$	$9{\cdot}6 \times 10^3$
2×10^{-5}	$1{\cdot}4 \times 10^5$	$6{\cdot}0 \times 10^4$
10^{-5}	$1{\cdot}4 \times 10^6$	$2{\cdot}4 \times 10^5$
5×10^{-6}	$2{\cdot}0 \times 10^7$	$9{\cdot}6 \times 10^5$

From these calculations, we may infer that it is unwise to attempt to perform experiments to measure directly 'equilibrium' distances of separation between parallel plates, since the attainment of any equilibrium position will be delayed by the viscous and electroviscous resistance to approach, which attain considerable dimensions at comparatively large distances of separation. Also, under these conditions, approach may be so slow that a false impression of equilibrium is given. For these reasons, the experiments of Buzagh (1937) would appear to be unsatisfactory.

The use of spherical surfaces would appear to be more suitable in the experimental study of this problem, since their attainment of equilibrium would be more rapid. Bubbles as used by Derjaguin & Kussakov, however, are not satisfactory, as flattening of the approaching surfaces leads to a slow rate of approach (see § 5).

5. EVIDENCE FOR THE EXISTENCE OF THE ELECTROVISCOUS EFFECT

(a) *Anomalous viscosity in narrow channels*

Terzaghi (1931) found experimentally that the apparent viscosity of water was increased in narrow channels. Macauley (1936) measured the viscosity of water between two parallel glass plates separated by a distance of $2{\cdot}5 \times 10^{-5}$ cm., and found a value of 0·11 poise compared with 0·01 poise in wide tubes.

Now by equation (63)

$$\eta_a = \eta + \frac{3\epsilon^2\zeta^2}{32\pi^2h^2\kappa} \text{ approx.,}$$

where $2h$ is the distance of separation of the plates.

If we take likely values for the above constants, η_a, the apparent viscosity, may be calculated. Putting $\epsilon = 80$, $\zeta = 150$ mV, and $\kappa = 10^{-6}$ ohm^{-1}cm.$^{-1}$, we arrive at $\eta_a = 0{\cdot}12$ poise, a value in good agreement with that found experimentally by Macauley.

Reekie & Aird (1945) found that the apparent viscosity of water in diaphragms of very small particle size (of the order 10^{-4} cm.) was 5 to 10 times the normal bulk viscosity.

(b) *Anomalous ζ-potentials in narrow capillaries*

Since in a narrow capillary tube

$$E = \frac{\zeta\epsilon P}{4\pi\eta_a\kappa}, \tag{66}$$

then if instead of the apparent viscosity, the normal bulk viscosity is used to calculate the electrokinetic potential ζ from the streaming potential E, an anomalous result ζ_a will be obtained such that

$$\frac{\zeta_a}{\zeta} = \frac{\eta}{\eta_a}. \tag{67}$$

In narrow capillaries, therefore, the apparent ζ-potential will be reduced if the normal bulk viscosity is used in the calculation.

From equations (28) and (67), we obtain

$$\zeta_a = \frac{3\pi^2a^2\eta\kappa\zeta}{3\pi^2a^2\eta\kappa + \zeta^2\epsilon^2}. \tag{68}$$

We would therefore expect the apparent electrokinetic potential to vary with the capillary radius as shown in figure 1.

Results in qualitative agreement with these calculations have been obtained by Bull & Gortner (1932), who measured apparent ζ-potentials in diaphragms of quartz particles of various sizes, and found that the apparent ζ-potential was reduced in diaphragms of very small particle size. In figure 2 their results for E/P (proportional to ζ_a) are plotted as a function of the particle size of the diaphragm, a curve of similar form to that given in figure 1 being obtained. Other workers (White, Urban & Krick 1932; Bishop, Urban & White 1931; White, Urban & van Atta 1932) have reported the observation of similar effects in narrow capillaries.

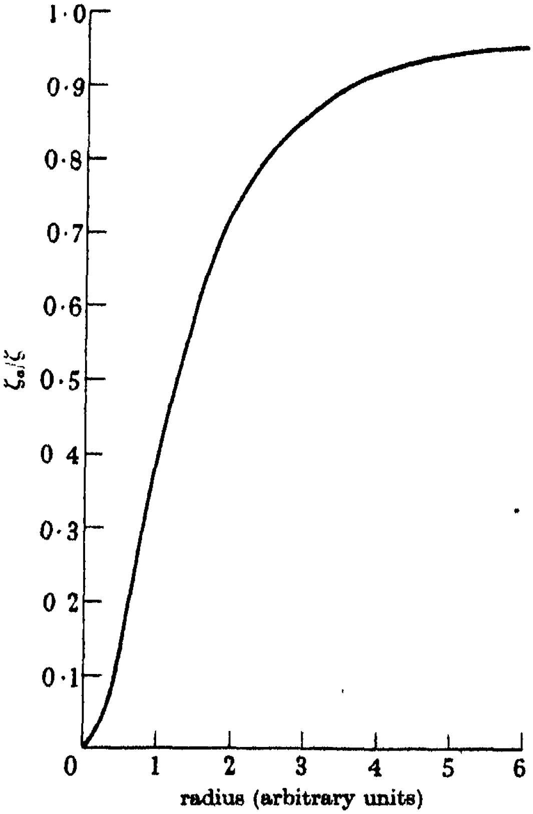


Figure 1. Expected variation of apparent ζ-potential with tube radius.

Bull & Meyer (1936) attempt to explain these results by allowing for a decrease in the effective radius of the tube due to the thickness of the double layer, but their calculations were of an empirical nature. Reichardt (1933) also made general theoretical observations on the effect. His calculations led him to expect the presence of a minimum in the curve relating apparent ζ-potential with the capillary radius. This has not been observed experimentally. It is submitted that the calculations given in the present paper lead to a satisfactory explanation of the observed phenomenon.

(c) *Surface conductivity*

Various workers (Fairbrother & Balkin 1931; Rutgers 1940; Bikermann 1942; McBain, Peaker & King 1929) have reported that the apparent conductivity of an ionic solution is increased in narrow capillaries. This increase, usually ascribed to 'surface conductivity', may, in the opinion of the writer, be capable of interpretation on a similar basis to the increase in viscosity, since in § 4 the conditions inside a tube containing ions have been completely defined.

It is proposed to make the detailed investigation of this effect the subject of a further publication.

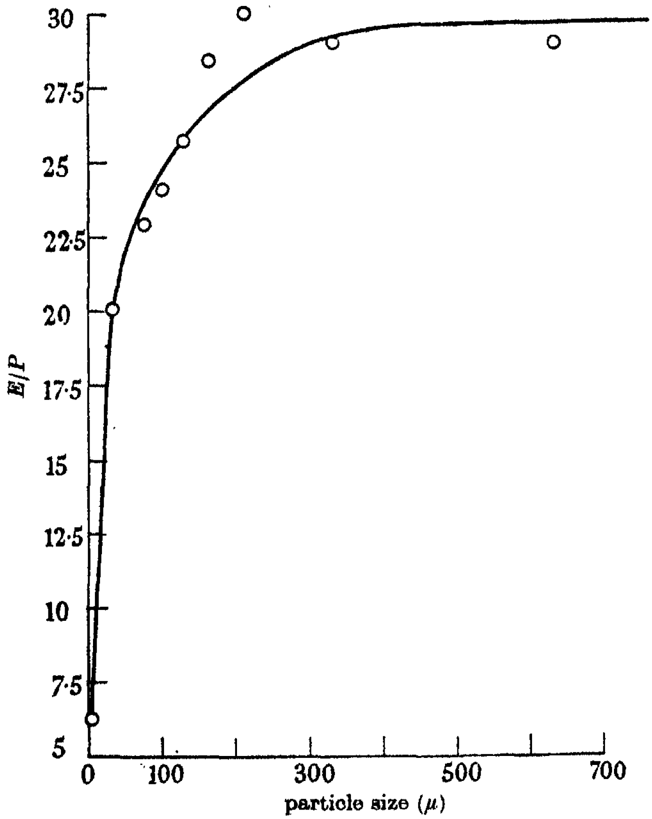


FIGURE 2. Variation of E/P with particle size in quartz diaphragms (Bull & Gortner: 2×10^{-4} N-NaCl).

(d) *The stability of lyophobic colloids*

In a lyophobic colloidal solution the rate of approach of two large particles under the influence of Brownian forces will depend on the total viscous and electroviscous resistance to flow of the medium between them; other conditions being equal, the greater the resistance to flow, the greater will be the stability of the solution. The total resistance to flow will vary with the apparent viscosity, which is a function of the ζ-potential of the particles, and the dielectric constant and specific conductivity of the medium. Addition of electrolytes, which will decrease the ζ-potential of the particles, and increase the conductivity of the medium, will decrease the total resistance to approach, and, consequently, the stability of the solution. Multivalent ions, which reduce the electrokinetic potential to a greater extent then univalent ions, will be more effective in reducing the stability of the colloid. At the isoelectric point the stability will be at a minimum ($\zeta = 0$, so that $\eta = \eta_a$). On reversing the sign of the electrokinetic potential by further addition of electrolyte, the stability will increase, but it is unlikely to reach its original value, because of the increase in the specific conductivity of the medium.

This mechanism would thus seem to offer a new qualitative explanation of the coagulation of lyophobic colloids by electrolytes, and it is intended to develop these ideas quantitatively in a future publication. It must be emphasized, however, that ordinary statistical treatment cannot be applied to a particle in motion under the action of Brownian (i.e. non-statistical) forces.

(*e*) *Approach of a fluid bubble to a surface*

Derjaguin & Kussakov (1939*a*) showed that the profile of a bubble approaching a surface was as shown diagrammatically in figure 3. The thinnest portion of the liquid film between the bubble and the surface is not in the centre, but in an annular ring ('barrier'), due to the more rapid escape of the liquid around the periphery. Derjaguin & Kussakov derive an expression (applicable only to small bubbles) for the radius ρ of this barrier ring:

$$\rho = r^2\left(\frac{2dg}{3\gamma}\right)^{\frac{1}{2}}, \tag{69}$$

where r is the radius of the bubble, d the difference in density between the bubble and the bulk liquid, g the gravitational constant, and γ the interfacial tension.

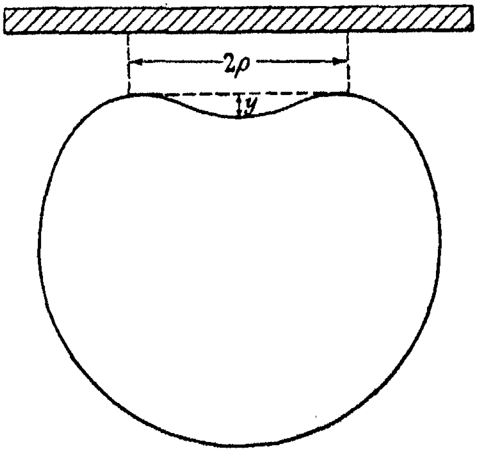


FIGURE 3

This distortion of the bubble will produce a diminution in the rate of approach. We may obtain some idea of the order of the effect (neglecting electroviscous forces) by calculating the barrier ring radius from equation (69), and then treating the barrier ring as a flat plate of equal radius, approaching the surface. This treatment is permissible to a first approximation, as the distance y (figure 3) may be shown to be very small compared with ρ.

Thus from equations (1) and (69), putting $m = \frac{4}{3}\pi r^3 d$, we have

$$t = \frac{r^5 dg}{4\gamma^2}\left(\frac{1}{h_2^2} - \frac{1}{h_1^2}\right), \tag{70}$$

and if $h_1 \gg h_2$

$$t = \frac{r^5 dg}{4\gamma^2 h_2^2}. \tag{71}$$

For a bubble of radius 0·1 cm. approaching a plane surface in water, equation (71) leads to a distance of separation of 1200 A after 1 hr., a value of the same order as the 'equilibrium' distances found by Derjaguin & Kussakov. If the calculated distance of approach at any one time is plotted against the reciprocal of the bubble radius (proportional to the pressure inside the bubble), a curve of the type shown in figure 4 is obtained.

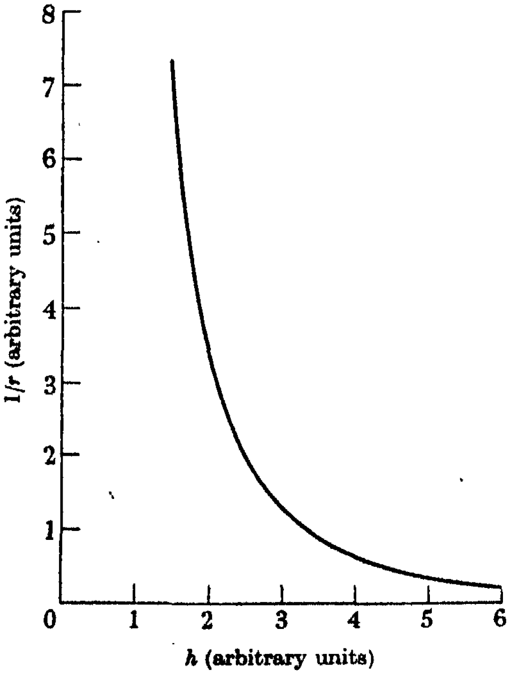


FIGURE 4

Curves very similar to this are obtained experimentally by Derjaguin & Kussakov. The dependence of the distance of separation on the ionic concentration (as observed experimentally) would be explained by the existence of a greater or lesser electroviscous resistance superimposed on the ordinary viscous resistance to approach. Experimental work performed by the writer (see next paper, p. 281) has served to confirm these conclusions.

6. CONCLUSIONS

1. It has been demonstrated that there is considerable theoretical and experimental basis for the belief that the electrical double layer at an interface in an ionic liquid can exhibit considerable resistance to shear.

2. Knowledge of the existence of such an electroviscous effect may be of considerable value in the explanation of hitherto unexplained phenomena.

3. The attainment of equilibrium distances of separation between surfaces may be considerably delayed by a combination of viscous and electroviscous forces preventing flow of the liquid between the surfaces. Some previous experimental work may be explicable without the postulate of long-range repulsive forces, or of rigid or compressible liquid films between surfaces.

REFERENCES

Bikermann 1942 *J. Phys. Chem.* **46**, 724.
Bishop, Urban & White 1931 *J. Phys. Chem.* **35**, 137.
Bull & Gortner 1932 *J. Phys. Chem.* **36**, 111.
Bull & Meyer 1936 *J. Phys. Chem.* **40**, 9.
Buzagh 1937 *Colloid systems.* London. Technical Press.
Derjaguin 1940 *Trans. Faraday Soc.* **36**, 203.
Derjaguin & Kussakov 1939*a* *Acta Physicochim. U.R.S.S.* **10**, 25.
Derjaguin & Kussakov 1939*b* *Acta Physicochim. U.R.S.S.* **10**, 153.
Elton 1946 *J. Chem. Phys.* **14**, 363.
Eversole & Lahr 1941*a* *J. Chem. Phys.* **9**, 530.
Eversole & Lahr 1941*b* *J. Chem. Phys.* **9**, 686.
Fairbrother & Balkin 1931 *J. Chem. Soc.* p. 1564.
Hardy & Bircumshaw 1925 *Proc. Roy. Soc.* A, **108**, 12.
Langmuir 1939 *J. Chem. Phys.* **6**, 873.
Levine & Dube 1939 *Trans. Faraday Soc.* **35**, 1125.
Macauley 1936 *Nature*, **138**, 587.
McBain, Peaker & King 1929 *J. Amer. Chem. Soc.* **51**, 3294.
Reekie & Aird 1945 *Nature*, **156**, 367.
Reichardt 1933 *Z. Phys. Chem.* A, **166**, 433.
Reynolds 1886 *Phil. Trans.* **177**, 157.
Rutgers 1940 *Trans. Faraday Soc.* **36**, 69.
Schofield 1946 *Faraday Soc. Symp. on Swelling and Shrinking.*
Terzaghi 1931 *J. Rheology*, **2**, 253.
Verwey & Overbeek 1946 *Faraday Soc. Symp. on Swelling and Shrinking.*
White, Urban & Krick 1932 *J. Phys. Chem.* **36**, 120.
White, Urban & van Atta 1932 *J. Phys. Chem.* **36**, 3152.

Electroviscosity

II. Experimental demonstration of the electroviscous effect

By G. A. H. Elton, *Chemistry Department, Battersea Polytechnic*

(*Communicated by J. Kenyon, F.R.S.—Received* 20 *December* 1947)

Preliminary experimental evidence for the existence of the electroviscous in ionic liquids is submitted. The theoretical predictions of the previous paper have been qualitatively confirmed, and quantitative experimental work is now in progress. The rate of approach of surfaces in various liquids has been observed, and it has been demonstrated that equilibrium distances of separation under finite pressures are very small.

1. Introduction

It is the object of this paper to give details of preliminary experimental work carried out to demonstrate the existence of an electroviscous effect in ionic liquids, as predicted theoretically in the preceding paper. This work has also thrown some light on the problem of the possible existence of long-range repulsive forces between surfaces immersed in a liquid.

As already pointed out on pp. 259, 273, the experiments of Eversole & Lahr, and Derjaguin & Kussakov, which appeared to support the existence of these forces, are open to serious criticism on theoretical grounds.

In the re-examination of this problem by the writer, it was proposed:

(i) To repeat the experiments of Eversole & Lahr (1941), using an apparatus and experimental technique capable of giving results of sufficient accuracy and precision to enable the drawing of definite conclusions on the presence or absence of thick rigid water layers at quartz/water interfaces, and to extend the work to suitable non-polar liquids.

(ii) To study the approach of a quartz lens to a quartz plate in water, in dilute salt solutions, and in non-ionic liquids, in an attempt to obtain experimental evidence of the electroviscous effect, and to check the results of the experiments proposed in (i) above.

(iii) To repeat and extend the experiments of Derjaguin & Kussakov (1939*a*, *b*) and Darjaguin (1940) on the behaviour of gas bubbles in close proximity to a liquid interface, in an attempt to obtain definite evidence of the presence or absence of high equilibrium distances of separation, and of the existence or non-existence of thick compressible solvate layers at these interfaces.

2. Condensation of a liquid film between a lens and plate initially in contact

Eversole & Lahr reported that if a water film were condensed between a quartz plate and convex lens, initially in contact, a rigid water film formed between the surfaces produced a measurable equilibrium distance of separation. Their results were criticized on statistical grounds by the writer (Elton 1946).

(i) *Materials used*

Two optically polished quartz plates, and a plano-convex quartz lens of radius of curvature 100 cm., were used in this investigation. All were grade A optical quality supplied by Messrs Thermal Syndicate Ltd.

The water used was conductivity water which had been allowed to come to equilibrium with the atmosphere, and had a specific conductivity of about $0{\cdot}7 \times 10^{-6}$ ohm^{-1}cm.$^{-1}$.

Methyl alcohol was purified according to the method of Weissburger & Proskauer (1935), and stored in a stoppered bottle in a desiccator containing magnesium perchlorate.

Toluene was purified according to the method described by Mathews (1926), and stored as described above.

Ether was purified by keeping the B.D.H. AnalaR material over sodium for a prolonged period, and distilling through a 6 ft. column packed with ¼ in. Lessing rings into a flask fitted with a magnesium perchlorate tube, first and last runnings being rejected. The resultant liquid was stored as described above.

Potassium chloride was prepared pure by recrystallizing B.D.H. AnalaR potassium chloride from conductivity water.

(ii) *Apparatus*

The form of the apparatus used in this investigation is shown in figure 1. The cell *C* consisted of a stout phosphor-bronze cylinder covered by a thick brass plate *P* standing on a square-sectioned rubber gasket *G* around the circumference of the cylinder. An optical glass observation window *W* was cemented into the centre of the brass plate, and the cell was rendered gastight by screwing down an annular cap *A* over the brass plate on to the cylinder. A brass tube bearing two leads to the remainder of the apparatus (figure 2) was let into the base of the cylinder. The cell itself stood on a horizontal metal table, which could be moved up and down a vertical column supported on a heavy base.

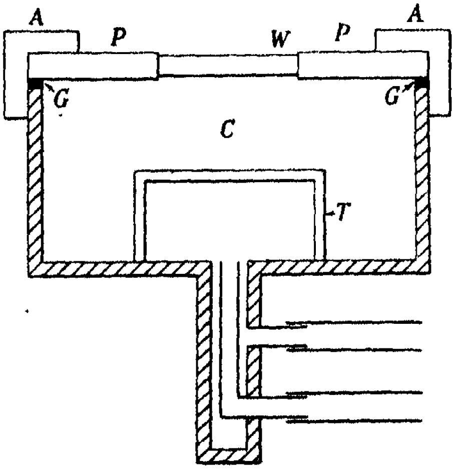


Figure 1

A small brass table *T* stood inside the cell, and on this were placed (*a*) a ground cobalt glass plate to absorb unwanted light reflexions, (*b*) the optically polished quartz plate, and (*c*) the optically polished quartz lens.

Monochromatic light was provided by a sodium vapour lamp placed behind a slit at the focus of a short focal length convex lens. The parallel beam was reflected down into the cell by a microscope cover-slip cemented to a rod which could be rotated about its axis and adjusted so that the cover-slip was at an angle of 45° to the vertical. The microscope was mounted in a brass collar which could slide up and down the supporting column, and could be locked in any desired position. A half-plate camera body was attached to the microscope, and the ground-glass focusing screen of the camera could be racked up and down independently, enabling very sharp focus to be obtained. Usually no microscope eyepiece was used, the image being projected directly on to the screen by the objective. Either a 4 in. or a 2 in. objective was used, according to the magnification required. Dry, dust-free air, free from carbon dioxide, could be drawn through the apparatus via a purification train.

(iii) *Experimental technique—preliminary testing*

When attempting to measure small distances of separation, scrupulous cleaning of the solid surfaces was essential. The quartz surfaces were vapour degreased using benzene vapour, traces of which were then removed by heating in an enclosure at 100 to 150° C. The clean dry lens and plate were placed in position on the brass table inside the enclosure (with the lens on top of the plate), the brass cover-plate replaced on the cell, and the cap screwed down. Clean dry air was drawn through the enclosure for some minutes, while the illumination system was adjusted. The Newton's rings formed between the surfaces were photographed on anti-halo backed Kodak O 800 photographic plates. Measurement of the radius of the rings was made from a positive print, which was mounted on a light-box and viewed with a low-power travelling microscope. Since the rings were of finite width, the inside and outside radii were measured and the average taken.

The distance of separation between the lens and plate was calculated as follows. It may be shown that

$$r_n^2 = \frac{Rn\lambda}{\mu} - 2Rt, \tag{1}$$

where r_n is the radius of the nth dark ring, R is the radius of curvature of the lens, λ is the wave-length of the light, μ is the refractive index of the medium between the surfaces, and t is the minimum distance of separation of the surfaces. A graph of r_n^2 against n gives a straight line having an intercept on the r_n^2 axis of $-2Rt$ and slope $\lambda R/\mu$. Since λ, R and μ are known, the slope may be calculated, and, if the rings had been measured directly, would be equal to the observed slope. If, however, the rings have been photographed, under conditions which render the exact determination of the magnification in each case difficult, the magnification M may be calculated as follows. If the values of the radius as measured on the photograph squared (r_1^2) are plotted against n, we have

$$r_1^2 = M^2 r_n^2 = \frac{M^2 n\lambda R}{\mu} - 2RtM^2. \tag{2}$$

If the observed slope is divided by the calculated slope, M^2 is obtained.

The observed values of r_1^2 were plotted against n, and the best line through them and the intercept on the r_1^2 axis calculated by the method of least squares. The magnification M was calculated as described above, and the calculated intercept on the r_1^2 axis was divided by $-2RM^2$ to give the required value of t.

TABLE 1

experiment	distance of separation (A)	experiment	distance of separation (A)
1	−20	7	−10
2	−20	8	0
3	+15	9	+5
4	+5	10	+20
5	0		
6	+20	mean	+2

TABLE 2. EXPERIMENTAL RESULTS OF EVERSOLE & LAHR

experiment	distance of separation in air, t_a (A)	distance of separation in water, t_w (A)	$t_w - t_a$ (A)
1	−1200	−870	330
2	−950	−800	160
3	−90	+140	230
4	−1160	−840	320
5	−350	−220	130
mean	-75_2	-51_8	32_4

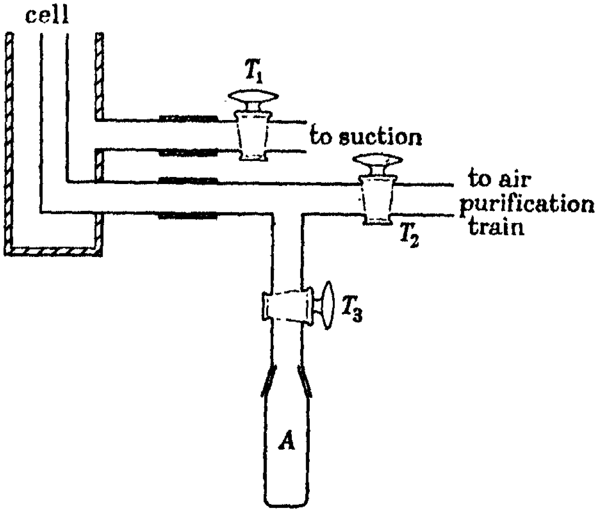


FIGURE 2

A large number of preliminary experiments were made to develop and test the experimental technique. Table 1 gives some of the final results obtained for the distance of separation of the lens and plate in air. Individual values of 0 ± 20 A were obtained.

The possible error of ± 20 A approx. compared very favourably with the errors of ± 1000 A approx. recorded by Eversole & Lahr (see table 2).

(iv) *Experiments on liquids*

Briefly, the procedure in these experiments was to place the lens and plate together in air, measure the distance of separation in air (t_{air}), introduce the liquid, measure the distance of separation in the liquid (t_{liquid}), evaporate off the liquid, and remeasure t_{air}. After several experiments, the results were examined statistically in order to detect any significant difference between t_{liquid} and t_{air} which might be taken as indicating the presence of a rigid liquid layer separating the two surfaces.

TABLE 3

liquid	experiment	t_{air} (A)	$t_{\text{liq.}}$ (A)	$t_{\text{liq.}} - t_{\text{air}}$ (A)	P (%)
water	1	20	15	− 5	
	2	− 5	5	10	
	3	− 5	0	5	
	4	10	0	− 10	
	5	0	5	5	
	mean	4	5	1	98
methyl alcohol	1	5	− 5	− 10	
	2	25	− 5	− 30	
	3	10	10	0	
	4	− 10	0	10	
	5	− 10	− 5	5	
	mean	4	− 1	− 5	75
toluene	1	− 10	− 10	0	
	2	0	− 5	− 5	
	3	0	20	20	
	4	15	5	− 10	
	5	15	20	5	
	mean	4	6	2	95
ether	1	5	25	20	
	2	− 15	− 20	− 5	
	3	20	− 5	− 25	
	4	10	10	0	
	5	0	10	10	
	mean	20	20	0	100

The lens and plate were placed together in the enclosure and the rings formed in air were photographed. The vapour of the liquid to be examined was then introduced using the apparatus shown in figure 2. The liquid was held in the container A, which was provided with a ground-glass neck for easy removal. The taps T_1 and T_2 were closed, tap T_3 was opened, and the liquid in A was warmed. Liquid condensed in the main enclosure between the lens and plate and the rings formed in the liquid were photographed. Tap T_3 was then closed, taps T_1 and T_2 opened, and a stream of pure dry air was drawn through the apparatus until the liquid ring had disappeared, and continued for a further considerable period of time. In the removal of this liquid ring, a persistent central spot of liquid always occurred, due to the lowering of the effective vapour pressure by the Kelvin effect in the narrow annular capillary formed between the lens and plate. (The quantitative study of this effect is to be the subject of a further publication.) Prolonged aspiration was necessary in order to complete the removal of this spot. When the aspiration was complete, the rings formed in air was rephotographed.

Measurements were made using water, methyl alcohol, toluene and ether. The results obtained are given in table 3. It is seen that in all cases there is no significant difference between the values of t_{liquid} and t_{air}. P in table 3 represents the probability of insignificance of the difference between the means of the two series.

3. Approach of a lens to a plate in a liquid

(i) *Experimental method*

A glass dish of suitable dimensions was placed inside the observation cell described in the previous section (figure 1), and the lens placed convex side downwards on the plate, which rested on the plate of ground cobalt glass. The liquid under examination, which had been prepared as free from dust as possible, was placed in the dish to a depth of about 2 cm. Using clean tongs, the lens was raised to a height of about 1 cm. above the plate and allowed to fall. The brass cover-plate was replaced on the cell, the cap screwed down, and the Newton's rings formed between the lens and the plate were photographed at suitable intervals. It was found that the rings obtained were not so clear and sharp as those obtained with air between the plates, due to the smaller difference in refractive index between the two media. Also, some undesirable reflexion of light occurred at the liquid/air interface. This could not be readily eliminated, but was not sufficiently serious to affect materially the results obtained.

(ii) *Strong salt solutions*

Solutions of potassium chloride in water were made up having strengths 1·0, 0·5, 0·1 and 0·05 N, and used in experiments as described in the preceding paragraph. It was found in these and subsequent experiments with liquids that the probable error was rather higher than before, probably about ± 50 A, due to the reduced sharpness of the rings. In all cases it is found that within the limits of experimental error, the lens has reached the plate before the first reading has been taken (usually less than 5 min.).

This is to be expected from ordinary viscous laws, in the absence of any appreciable electroviscous effect. The lens would presumably approach rather faster than a sphere of the same radius of curvature acted upon by the same force, since the viscous drag would be less. Calculations from the Taylor equation on this basis lead to very small distances of approach within 1 min.

(iii) *Non-ionic liquids*

The liquids used were methyl alcohol, toluene and ether. Great care was taken to keep the liquids absolutely dry during storage, and during the experiments a few crystals of magnesium perchlorate were placed in the liquids.

Under these conditions, it was again found to be impossible to detect any motion of the lens after a few minutes. If, however, the liquids contained even a trace of water, the lens approached more slowly. Table 4 gives the experimental results obtained using benzene containing 0·1 % added conductivity water. Measurable approach is continuing after more than 24 hr.

TABLE 4. APPROACH OF LENS TO PLATE IN BENZENE CONTAINING 0·1 % ADDED CONDUCTIVITY WATER

time (sec.)	distance of separation (A)
900	450
3,200	340
9,300	250
10,300	200
88,800	120
94,200	100

This phenomenon is attributed to the existence of a finite electrokinetic potential against the quartz in the case of these mixtures. This, together with the extremely low conductivities of the mixtures, would be expected from the previous paper to lead to an appreciable resistance to approach. (*Note*. For zero electrokinetic potential and zero specific conductivity, the apparent viscosity becomes indeterminate (see p. 267). However, under these conditions, since no ions are present, the electroviscous resistance to flow is zero, so that $\eta = \eta_a$).

(iv) *Dilute salt solutions*

It is in very dilute salt solutions and in pure water that the greatest electroviscous forces are to be expected, on account of the high electrokinetic potentials and low conductivities of these liquids.

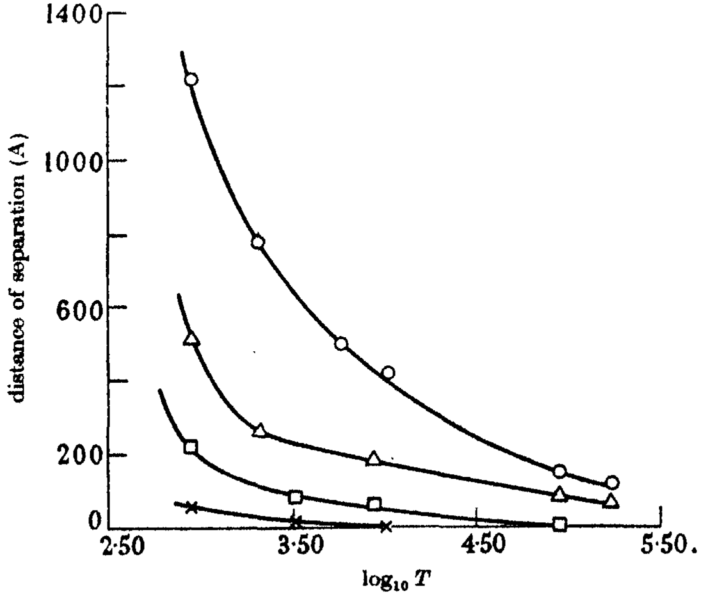


FIGURE 3. Distances of approach of lens in dilute salt solutions. ○ water, △ 10^{-5} N-KCl, □ 10^{-4} N-KCl and × 10^{-3} N-KCl.

Solutions of potassium chloride in conductivity water were prepared having strengths 10^{-5}, 10^{-4}, 10^{-3} and 10^{-2} N. The approach of the lens to the plate in these solutions was observed, and the results obtained are expressed graphically in figure 3.

It is seen that for small distances of separation, the rate of approach of the lens *to the plate in these solutions is markedly dependent on the ionic concentration*, increasing as the ionic concentration increases.

4. THE APPROACH OF AN AIR BUBBLE TO A SOLID SURFACE IMMERSED IN A LIQUID

(i) *Experimental method*

The experimental arrangement was a simple one. An optically polished plano-concave quartz lens of radius of curvature 100 cm. was vapour degreased and then placed concave side down on three Pyrex glass blocks arranged tripod fashion over a ground cobalt glass plate in a Pyrex glass dish. (The concave quartz surface was used to overcome the difficulty originally encountered in maintaining a free air bubble motionless under a flat plate.)

When the concave lens had been placed in position, the liquid under examination was introduced into the dish so as to cover the lens completely, and a small air bubble released under the lens by means of a pipette. The dish was then covered with a flat glass plate, the optical system adjusted, and the Newton's rings formed in the bubble observed.

As a bubble approached the plate it was observed that the innermost rings moved towards the centre, while the outer rings moved away from the centre of the ring system. This behaviour indicated the presence of a 'barrier' ring (see p. 272 and Derjaguin & Kussakov 1939*a*), having a minimum distance of separation from the plate. The barrier ring was located between the inward-moving and the outward-moving Newton's rings.

When the motion of the bubble was sufficiently slow, the rings were photographed. The diameter of the rings was measured on the photograph and the profile was constructed. A typical photograph and the corresponding profile was given in figures 4 and 5.

The minimum distance of separation of the bubble from the plate could then be calculated as follows:

The length of a chord AB (figure 5) at a known height above the plate was measured on the profile, and the lengths of a number of such chords plotted against the corresponding heights. It was generally found that for heights up to about 10,000 A, a straight-line relationship was obtained, i.e.

$$l = A + Bh, \tag{3}$$

where l is the length of the chord, and h is the distance from the chord to the plate. In cases where experimental results deviated from this relationship, the equation

$$l = A + Bh + Ch^2 \tag{4}$$

was found to be satisfactory. In these equations A, B and C are constants. Use of one of these equations enabled the value of h corresponding to zero l to be calculated.

This value gave the minimum optical distance of separation of the bubble from the plate. It was therefore not necessary to know the exact magnification used in the photograph. The actual distance of separation could be obtained by dividing the optical distance of separation by the refractive index of the liquid.

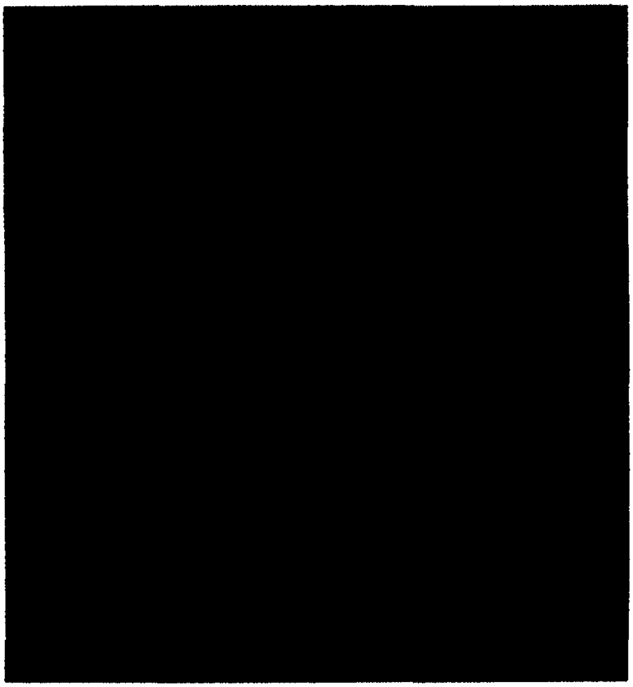


Figure 4

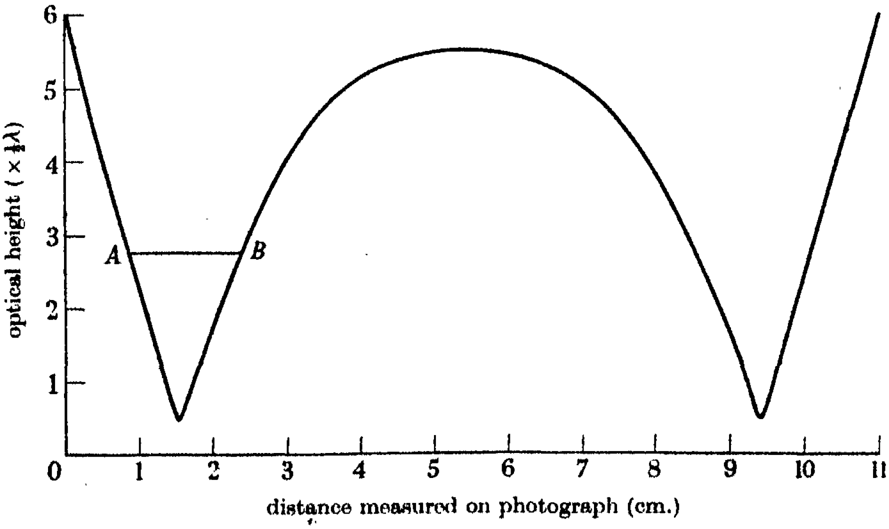
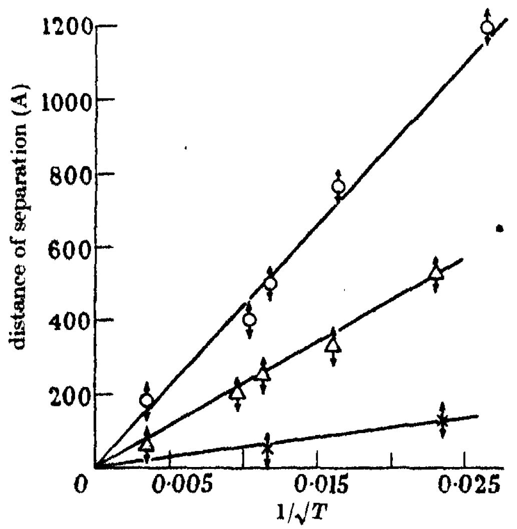


Figure 5. Bubble profile in water: distance of separation 1000 A.

Before discussing the results obtained, it is desirable to point out some of the experimental difficulties which were encountered during the course of the work:

(1) The distortion of the bubble from the spherical shape made necessary the devising of the special extrapolation method used for the calculation of the closest distance of approach. This will introduce a higher degree of uncertainty than the corresponding extrapolation for a spherical surface.



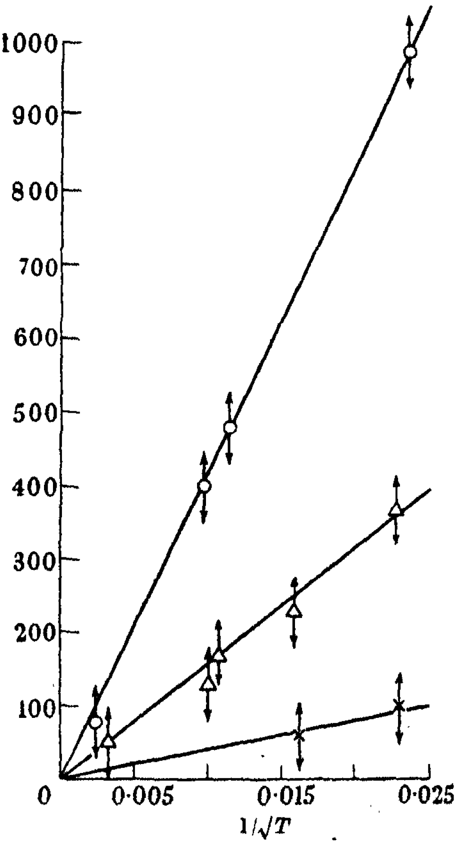


FIGURE 6. Distances of approach of bubble in benzene. ○ $r = 0{\cdot}129$ cm. △ $r = 0{\cdot}094$ cm. × $r = 0{\cdot}041$ cm.

FIGURE 7. Distance of approach of bubble in ether. ○ $r = 0{\cdot}125$ cm. △ $r = 0{\cdot}085$ cm. × $r = 0{\cdot}049$ cm.

(*Note*. The estimated experimental errors of ± 50 A have not, in the interest of clarity, been indicated on these graphs.)

(2) The small radius of curvature of the bubble gave a high rate of change of distance between the bubble/liquid and the solid/liquid interfaces. This means that the Newton rings are formed very close together, making necessary the use of higher magnification in the photographs (usually about 100 times), and involving a lower degree of accuracy in the location of any individual ring than is possible when the rings are widely spaced.

(3) The liquid under examination and the air in the bubble must be mutually saturated to ensure that no alteration in the size of the bubble takes place.

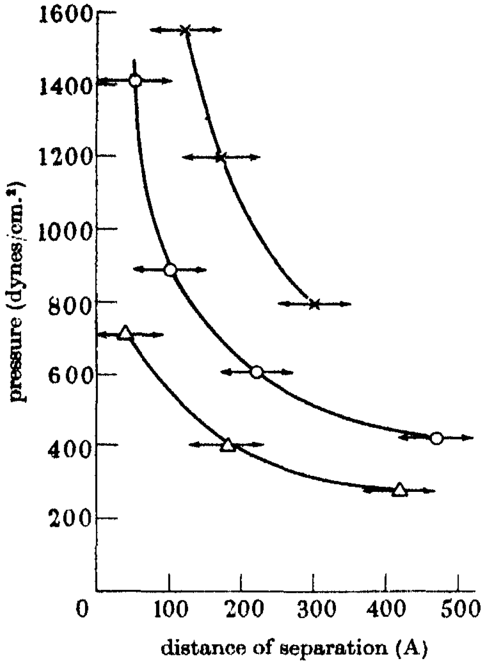


FIGURE 8. Bubble separation in benzene and ether after 10,000 sec. ○ benzene, △ ether and × 1 N-KCl.

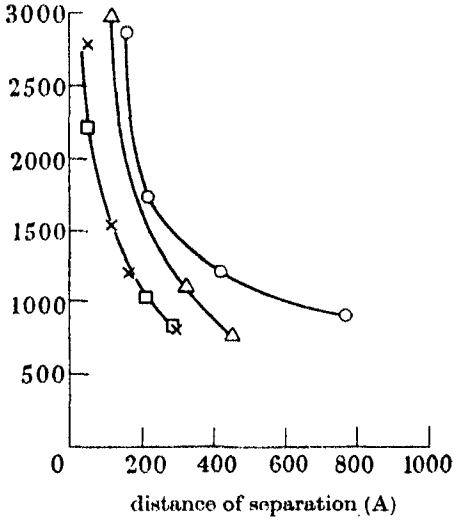


FIGURE 9. Distances of separation of bubbles in aqueous liquids after 10,000 sec. ○ water, △ 10^{-4} N-KCl, □ 10^{-2} N-KCl and × 1 N-KCl.

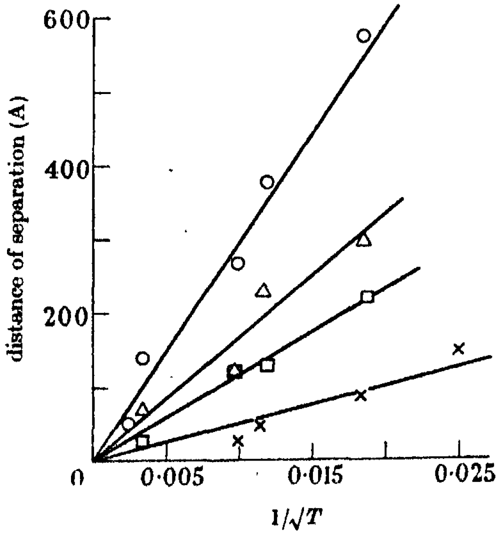


FIGURE 10. Approach of bubbles in 1 N potassium chloride solution. ○ $r = 0{\cdot}052$ cm., △ $r = 0{\cdot}094$ cm., □ $r = 0{\cdot}122$ cm. and × $r = 0{\cdot}182$ cm.

(*Note.* The estimated experimental errors of ± 50 A have not, in the interest of clarity, been indicated on these graphs.)

This requires accurate temperature control, which was not feasible in these experiments.

(4) Owing to distortion, the pressure inside the bubble, if required, is not easy to calculate.

Owing to these experimental difficulties, the results obtained were not of a high order of accuracy, but it is submitted that certain definite qualitative conclusions may be drawn from them. Figures 6 to 10 give distances and time of approach to the plate for bubbles of various sizes in dry benzene and ether, in water and dilute solutions of potassium chloride, and in concentrated potassium chloride solutions respectively. The possible experimental error is represented on the graphs as ± 50 A.

It is seen that with small bubbles (radius approx. 0·05 cm.), the rate of approach to quite small distances (less than 100 A) is fairly rapid in the non-ionic liquids and in the concentrated potassium chloride solutions. For larger bubbles the times of approach are considerably greater, measurable approach continuing after more than 48 hr. in the case of bubbles of radius approx. 0·125 cm.

Figures 6 and 7 demonstrate that, for bubbles of constant radius approaching the plate in the same non-ionic liquid, the distance of separation (when this is small) is approximately inversely proportional to the square root of the time of approach: this is to be expected from the approximate equation (71) in the preceding paper (p. 273).

In the experiments with water and dilute aqueous solutions of potassium chloride, it is seen that the distance of separation falls off less rapidly than in the non-ionic liquids, the distance now being approximately inversely proportional to the fourth root of the time of approach.

Figures 8 and 9 correlate graphically the distances of separation for bubbles (in the various liquids at a fixed time, 10,000 sec.) with the pressures inside them, calculated from the surface tension and the radius (neglecting distortion). It is seen that the curves so derived are of the same form as the 'equilibrium separation' curves given by Derjaguin & Kussakov. However, after this time slow but measurable approach is continuing, and equilibrium has not been reached.

Further, the rate of approach of a bubble of given radius is found to be dependent on the ionic concentration in the dilute potassium chloride solutions, decreasing as the ionic concentration decreases. This fact, together with the observation that the distance of approach is no longer inversely proportional to the square root of the time of approach is taken as indicating the operation of an electroviscous resistance to approach in dilute salt solutions, this resistance decreasing as the ionic concentration increases, as previously predicted (p. 273).

The results obtained in concentrated salt solutions (e.g. in 1 N-KCl, figure 10), where the distance of approach is found to be approximately inversely proportional to the square root of the time of approach (as with the non-ionic liquids) are thus also in agreement with the prediction. In concentrated salt solutions, where the electrokinetic potential is small, and the specific conductivity large, the electroviscous resistance to approach is negligible compared with the ordinary viscous resistance, even at very small distances of approach.

5. CONCLUSIONS

1. The work described in this paper has demonstrated that rigid solvate layers of thickness 100 A or more (as reported by Eversole & Lahr) do not exist at quartz interfaces in water, methyl alcohol, toluene and ether. If a rigid layer of ions or molecules is attached to the surface, it is of thickness considerably less than 25 A.

2. The high distances of separation between surfaces pressed together in liquids, observed by Derjaguin & Kussakov do not correspond to equilibrium states, but are the result of slow approach under the influence of viscous and electroviscous forces.

3. Qualitative evidence for the existence and importance of the electroviscous effect in ionic solutions has been obtained. Further experimental work is now in progress, with the object of testing directly the quantitative equations derived in the preceding paper.

REFERENCES

Derjaguin 1940 *Trans. Faraday Soc.* **36**, 203.
Derjaguin & Kussakov 1939*a* *Acta Physicochim. U.R.S.S.* **10**, 25.
Derjaguin & Kussakov 1939*b* *Acta Physicochim. U.R.S.S.* **10**, 153.
Elton 1946 *J. Chem. Phys.* **14**, 363.
Eversole & Lahr 1941 *J. Chem. Phys.* **9**, 686.
Mathews 1926 *J. Amer. Chem. Soc.* **48**, 562.
Weissburger & Proskauer 1935 *Organic solvents*, p. 114. London: Oxford University Press.

The use of flat-ended projectiles for determining dynamic yield stress

I. Theoretical considerations

BY SIR GEOFFREY TAYLOR, F.R.S.

(*Received* 24 *July* 1947)

It has long been known that metals may be subjected momentarily to stresses far exceeding their static yield stress without suffering plastic strain. One of the simplest methods for subjecting a metal to a high stress for a short time is to form it into a cylindrical specimen and fire this at a steel target. The front part of this projectile crumples up, but the rear part is left undeformed. If the target is rigid the distance which this portion travels while it is being brought to rest may be taken as the difference between the initial length and the length of the deformed specimen after impact. Knowing the velocity of impact, a minimum possible value can be assigned to the maximum acceleration of the material, and from this a minimum value for the yield stress can be calculated. The actual yield stress is considerably greater than this minimum, and methods are given for calculating a more probable value.

INTRODUCTION

When a cylindrical projectile strikes perpendicularly on a flat rigid target, the stress at the impact end immediately rises to the elastic limit, and an elastic compression wave travels towards the rear end. The stress in this wave is equal to the elastic limit. If the material is one in which the stress rises when the strain exceeds that corresponding with the elastic limit, the elastic wave is followed by a plastic one. On reaching the rear end of the projectile, the elastic wave is reflected as a wave of tension which is superposed on the compression wave. At this stage the velocity of the material in the part of the projectile which the reflected wave has not yet reached is $U - S/\rho c$, where S is the yield stress of the material, ρ its density, c the velocity of elastic waves, and U the velocity of impact. The stress in this portion is S. In the reflected wave extending from the wave front to the rear end of the projectile the velocity is $U - 2S/\rho c$ and the stress is zero.

The reflected elastic wave runs forward along the projectile until it meets the front of the plastic wave advancing from the target plate. In this plastic wave, the stress will not rise appreciably above the yield stress at any point close to the plastic-elastic boundary, but the velocity may be nearly zero, or, at any rate, will be very different from that in the rear part where plastic flow has not taken place. At the moment when the reflected wave has just reached the plastic boundary, the part of the specimen which lies behind this is stress-free, and is moving as a solid body with velocity $U - 2S/\rho c$. It is, therefore, in the same condition as the projectile at the moment of impact, except that its speed is $U - 2S/\rho c$ instead of U, and its length is less than the original length L.

The length x of this portion which has not yet suffered plastic strain will depend on the speed of the projectile, the speed of elastic waves in it, and the velocity with which the plastic-elastic boundary moves away from the target plate. Under the conditions of all the experiments which are here considered, c is much greater than U, the velocity of the projectile, or V, the velocity of the plastic-elastic boundary, and S is comparable with ρU^2, but small compared with ρc^2.

If h is the distance of the plastic boundary from the target plate at any time, x the length of the portion which has not yet been plastically compressed, and u the velocity of this rear portion, the above considerations lead to the following equations for the small changes in u, h and x during one passage of an elastic wave from the plastic boundary and back to it. The duration of this double passage is

$$dt = \frac{2x}{c}, \tag{1}$$

so that

$$dh = v\frac{2x}{c}, \tag{2}$$

$$dx = -(u+v)\frac{2x}{c} \tag{3}$$

$$du = -\frac{2S}{\rho c}. \tag{4}$$

Eliminating c, equations (1) to (4) reduce to

$$\frac{dh}{dt} = v, \tag{5}$$

$$\frac{dx}{dt} = -(u+v), \tag{6}$$

$$\frac{du}{dt} = -\frac{2S}{2x\rho} = -\frac{S}{\rho x}. \tag{7}$$

It will be noticed that (5), (6) and (7) are the equations which would be derived if the rear portions of the projectile were regarded as rigid, and all the quantities as continuously varying.

The equations (6) and (7) are not sufficient to determine the motion. In fact, the velocity of the plastic boundary, v, is determined by the plastic flow between this boundary and the target. To analyze the dynamics completely, it would be necessary to know all the intermediate states of the projectile between the instant of impact and the time when it comes to rest, or leaves the target plate. The object of the present work is to extract as much information as possible from measurements of the projectile recovered after impact. In the absence of measurements made during the impact, it is necessary to make some assumption about how the plastic boundary moves from the surface of the target to the final position in which it is measured after the projectile has come to rest.

SIMPLE THEORETICAL MODEL

In order to obtain a simplified picture of the phenomenon to serve as a framework for thinking about the motion, the simplest possible assumption about the plastic stress-strain relationship was made, namely, that the stress in the part of the projectile where the material is yielding is constant and equal to the yield stress S. Further, the radial inertia is neglected, so that the stress can be considered as constant over any cross-section. It is not possible to discuss the plastic flow between the plastic boundary and the target without analyzing the complete problem of plastic waves. It is possible, however, to imagine a state in which the material which has just passed through the plastic boundary is brought to rest in a very short length. For this to be possible, the material must spread out very rapidly. The appearance of the theoretical model at a time when the rear end is still moving is shown at the right-hand side of figure 1. If A_0 is the cross-section of the projectile before it has been compressed plastically, and A the area at the point where the material is brought to rest, the continuity equation is

$$A_0(u+v) = Av, \tag{8}$$

and if the stress is S on both sides of the thin region where the change in area occurs, the momentum equation is

$$\rho A_0(u+v)\,u = S(A-A_0). \tag{9}$$

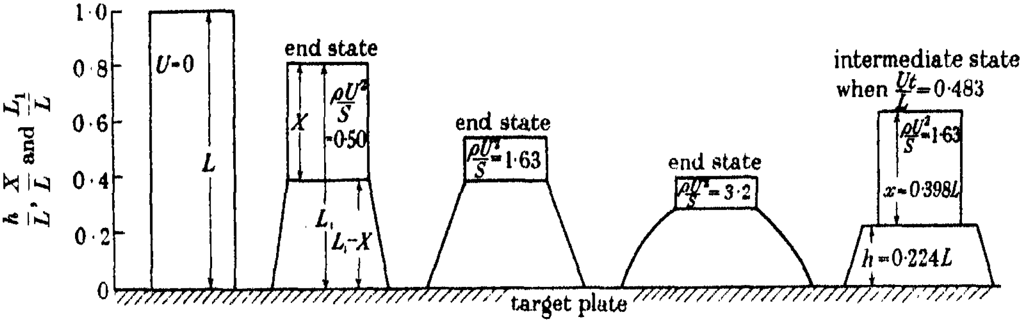

$\frac{h}{L}, \frac{X}{L}$ and $\frac{L_1}{L}$
1·0
0·8
0·6
0·4
0·2
0
$U=0$
L
end state
X
$\frac{\rho U^2}{S}=0{\cdot}50$
L_1
L_1-X
end state
$\frac{\rho U^2}{S}=1{\cdot}63$
end state
$\frac{\rho U^2}{S}=3{\cdot}2$
intermediate state
when $\frac{Ut}{L}=0{\cdot}483$
$\frac{\rho U^2}{S}=1{\cdot}63$
$x=0{\cdot}398L$
$h=0{\cdot}224L$
target plate

FIGURE 1. Simple theoretical model of flat-ended projectile fired at speed U at flat target.

The longitudinal compressive strain at any point may be defined as

$$e = 1-\frac{A_0}{A}. \tag{10}$$

Eliminating v from (8) and (9), and employing (10) to eliminate A and A_0, we obtain

$$\frac{\rho u^2}{S} = \frac{e^2}{1-e}. \tag{11}$$

From (6), (7), (8) and (10)

$$\frac{dx}{du} = \frac{(u+v)\rho x}{S} = \frac{\rho u x}{Se}. \tag{12}$$

Integrating (12)

$$\log_e(x^2) = \int \frac{1}{e}\,d\left(\frac{e^2}{1-e}\right) = \frac{1}{1-e} - \log_e(1-e) + \text{constant}. \tag{13}$$

At the moment of impact $x = L$, and $e = e_1$, say, and from (11)

$$\frac{\rho U^2}{S} = \frac{e_1^2}{1-e_1}. \tag{14}$$

When the projectile is brought to rest, $x = X$ and $e = 0$. X is one of the lengths which can be measured, hence

$$\log_e\left(\frac{x}{L}\right)^2 = \frac{1}{1-e} - \log_e(1-e) - \frac{1}{1-e_1} + \log_e(1-e_1) \tag{15}$$

and

$$\log_e\left(\frac{X}{L}\right)^2 = 1 - \frac{1}{1-e_1} + \log_e(1-e_1). \tag{16}$$

Eliminating e_1 between (14) and (16) gives X/L as a function of $\rho U^2/S$. The results are given in table 1 and shown graphically in figure 2.

TABLE 1. RESULTS OF CALCULATION BASED ON SIMPLE THEORETICAL MODEL

e_1	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
X/L	1·0	0·897	0·789	0·675	0·555	0·430	0·299	0·171	0·061	0·003	0
$\rho U^2/S$	0	0·011	0·050	0·128	0·267	0·500	0·900	1·633	3·200	8·10	—
h/L	—	—	—	—	—	0·382	—	0·376	0·288	—	—
L_1/L	—	—	—	—	—	0·812	—	0·547	0·349	—	—

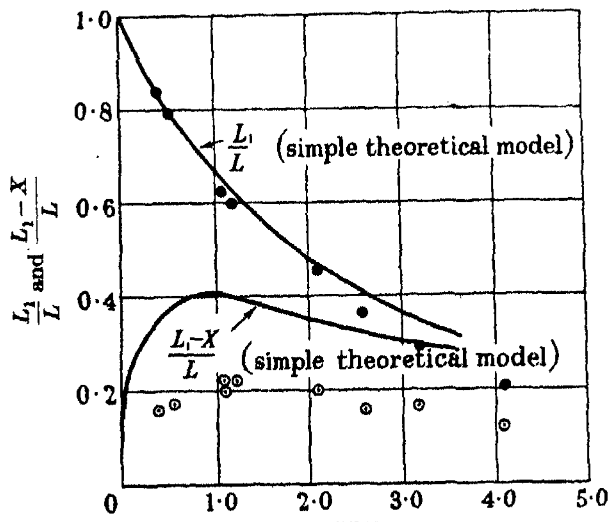


FIGURE 2. Results of calculation based on a simple theoretical model. ● measured values L_1/L; ⊙ measured values $(L_1-X)/L$.

To find the shape of the projectile after impact, we obtain from (5) and (6)

$$\frac{dh}{dx} = \frac{-v}{u+v} = -1+e,$$

so that

$$h = -\int_L^x (1-e)\,dx. \tag{17}$$

This integration was performed numerically in three cases, $e_1 = 0·5$, 0·7 and 0·8. Taking a given value of e_1, $\rho U^2/S$ was taken from table 1, and (x/L) calculated from (15) for the range of values of e from 0 to e_1. The resulting values of h for $e_1 = 0·5$ and

$e_1 = 0{\cdot}7$ are given in table 2. The corresponding cross-section is $A = A_0/(1-e)$, so that the diameter of the projectile at this point is $d = d_0\sqrt{\{1/(1-e)\}}$, where d_0 is the diameter of the projectile before firing. The values of d/d_0 and h/L are given in table 2. Similar calculations were made for $e_1 = 0{\cdot}8$.

TABLE 2. CALCULATIONS FOR SIMPLE THEORETICAL MODEL OF PROJECTILE FOR TWO IMPACT VELOCITIES

$e_1 = 0{\cdot}70, \quad \rho U^2/S = 1{\cdot}633$

e	0	0·05	0·10	0·15	0·20	0·25	0·30	0·35
x/L	0·171	0·180	0·190	0·202	0·216	0·233	0·253	0·277
h/L	0·376	0·368	0·358	0·348	0·336	0·323	0·309	0·293
Ut/L	0·720	0·708	0·695	0·681	0·665	0·646	0·625	0·599
d/d_0	1·00	1·026	1·054	1·087	1·118	1·155	1·196	1·241

e	0·40	0·45	0·50	0·55	0·60	0·65	0·70
x/L	0·307	0·346	0·398	0·468	0·571	0·730	1·000
h/L	0·274	0·251	0·224	0·191	0·147	0·088	0
Ut/L	0·568	0·531	0·483	0·420	0·333	0·207	0
d/d_0	1·292	1·349	1·414	1·492	1·583	1·690	1·827

$e_1 = 0{\cdot}50, \quad \rho U^2/S = 0{\cdot}50$

e	0	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50
x/L	0·430	0·452	0·478	0·508	0·543	0·585	0·635	0·696	0·773	0·871	1·000
h/L	0·382	0·361	0·337	0·311	0·282	0·250	0·213	0·172	0·124	0·067	0
Ut/L	0·332	0·315	0·298	0·277	0·255	0·229	0·199	0·163	0·119	0·067	0
d/d_0	1·000	1·026	1·054	1·087	1·118	1·155	1·196	1·241	1·292	1·349	1·414

The shapes of the projectile in the three cases are shown in figure 1 for the case where the diameter was initially 0·3 of the height. Comparing these with the profiles of the steel slugs shown in figure 1 of Mr Whiffin's paper (see part II), it will be seen that the calculated shape for $\rho U^2/S = 0{\cdot}5$ is very similar to that of the slug fired at 810 ft./sec. The shapes of the slugs fired at greater speeds do not resemble at all closely those calculated for $\rho U^2/S = 1{\cdot}63$ or $3{\cdot}2$. The plastically strained parts of the slugs fired at speeds greater than 810 ft./sec. have a concave profile, due, apparently, to the high radial velocity imparted to the material near the target which was neglected in the analysis.

As previously stated, the object of these calculations was to form the basis for an appropriate rough assumption for the rate at which the plastic boundary reaches its final position. To determine how h varies with t, the time since the beginning of the impact, equation (12) was integrated numerically. Using equations (7) and (11)

$$t = -\int \frac{\rho x}{S}\,du = -\int \frac{\rho x}{S}\sqrt{\frac{S}{\rho}}\,d\left(\frac{e}{\sqrt{(1-e)}}\right) = L\sqrt{\frac{\rho}{S}}\int \frac{x}{L}\frac{(1-\frac{1}{2}e)}{(1-e)^{\frac{3}{2}}}\,de,$$

and, using (14), this may be written

$$\frac{Ut}{L} = \frac{e_1}{\sqrt{(1-e_1)}}\int_e^{e_1} \frac{x}{L}\frac{(1-\frac{1}{2}e)}{(1-e)^{\frac{3}{2}}}\,de. \qquad (18)$$

The integration of (18) was performed numerically in the cases when $e_1 = 0{\cdot}5$ and $e_1 = 0{\cdot}7$. The results are given in table 2, together with the values of h. The relationship between Ut/L and h/L for these two values of e_1 are shown in figure 3. It will be seen that h increases nearly uniformly with Ut/L, so that the velocity of the plastic boundary, as it moves away from the target, is nearly uniform.

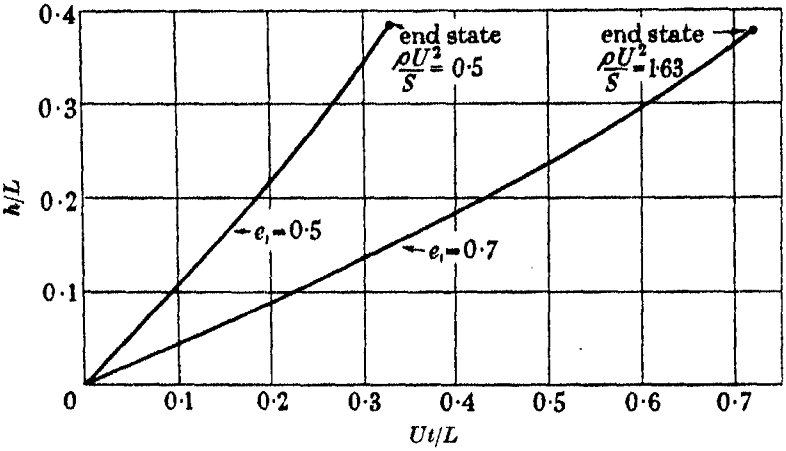


Figure 3. Propagation of plastic boundary in a simple theoretical model.

The complete system of plastic waves occurring during an impact has been analyzed by Messrs E. H. Lee and S. J. Tupper in a case where the plastic stress-strain relationship was assumed known. Good agreement was found with the strains calculated using the simple theoretical model here described. The main difference is that the strains calculated by the complete plastic-wave theory change in discontinuous jumps along the length of the projectile. This, as the authors point out in a paper which is not yet published, is due to the neglect in their analysis of the radial inertia of the plastic material as it spreads out near the target plate. The present calculations also neglect this radial inertia, but the discontinuities of strain disappear from the calculations when the finite difference equations (1), (2) and (3) are replaced by the differential equations (5), (6) and (7).

It is worth pointing out that though the plastic wave system can be calculated with considerable difficulty, when the plastic stress-strain relationship is known, this relationship can only be measured at low rates of straining. It is not possible to use these calculations directly to determine the stress-strain relationships from the shape of a projectile measured after an impact. It appears that the most which can be done at present is to determine approximately the yield stress using simplifying assumptions as to the general nature of the plastic stress waves. This is the line of attack developed in the present work.

APPROXIMATE FORMULA FOR ESTIMATING YIELD POINT FROM MEASUREMENTS OF SLUGS AFTER IMPACT

In developing a simple formula for estimating the yield point from measurements of the position of the yield boundary after the impact, it will be assumed that the plastic-elastic boundary moves outwards at a uniform velocity from the impact end to its final position. This, together with the assumption that the yield boundary represents the position where a definite compressive stress S is reached, is sufficient to fix the whole history of the deceleration of the rear part of the projectile.

If C is written for the constant velocity of the plastic boundary, equations (6) and (7) give

$$\frac{du}{dx} = \frac{S}{\rho x(u+C)}, \tag{19}$$

which may be integrated to give

$$\frac{S}{\rho}\log_e\left(\frac{x}{L}\right) = \tfrac{1}{2}u^2 + Cu - \tfrac{1}{2}U^2 - CU. \tag{20}$$

When $u = 0$, $x = X$, so that

$$\frac{S}{\rho}\log_e\left(\frac{X}{L}\right) = -\tfrac{1}{2}U^2 - CU. \tag{21}$$

If the deceleration of the rear of the projectile were exactly uniform, C/U could be determined simply from the fact that the time of deceleration T is equal to $(L_1 - X)/C$ (where L_1 = overall length of the projectile after test), and is also equal to $2(L - L_1)/U$, so that $C/U = \frac{1}{2}(L_1 - X)/(L - L_1)$, and equation (21) would then assume the form

$$\frac{S_1}{\rho U^2} = \frac{(L-X)}{2(L-L_1)}\frac{1}{\log_e(L/X)}, \tag{22}$$

where S_1 has been used instead of S to distinguish it from the value calculated by more exact methods. In fact, the deceleration is not uniform, so that (22) is only approximate. It is, however, the formula used with success by Mr Whiffin to interpret his experiments (see part II). In Mr Whiffin's experiments, slugs fired at very different speeds gave different values of X/L and L_1/L, but, when the measured values were inserted in (22), the values of S so deduced were found to be nearly independent of U. This fact affords a strong confirmation that (22) is an approximate equation which can be relied on to give the values of S from measurements of U, X/L and L_1/L. In practice, it is sometimes found that the slug makes a depression in the target. In this case, equation (22) can still be applied, but if L_1 is the total measured final length, and d the depth of the hole into which it fitted at the end of the impact, L_1 in (22) must be replaced by $(L_1 - d)$.

MORE EXACT CALCULATION

The fact that the deceleration of the rear of the projectile is not uniform under the conditions assumed introduces an error which can be calculated. Since

$$\frac{dx}{dt} = -(u+C),$$

equation (20) may be written

$$\left(\frac{dx}{dt}\right)^2 = \frac{2S}{\rho}\log_e\left(\frac{x}{L}\right) + (U+C)^2, \tag{23}$$

and when $u = 0$, $x = X$, so that

$$\frac{2S}{\rho}\log_e\frac{X}{L} = C^2 - (U+C)^2. \tag{24}$$

Writing $2S/\rho = a^2$, $K = (U+C)/a$, $x_1 = x/L$, $t_1 = at/L$, and $T_1 = aT/L$, where T is the duration of the impact, the non-dimensional form of (23) is obtained:

$$\frac{dx_1}{dt_1} = \sqrt{(K^2 + \log_e x_1)},$$

so that

$$T_1 = \int_1^{e^{-K^2}} \frac{dx_1}{\sqrt{(K^2+\log_e x_1)}}, \tag{25}$$

and, putting $K^2 + \log_e x_1 = z^2$,

$$T_1 = 2e^{-K^2} \int_{C/a}^{K} e^{z^2}\, dz. \tag{26}$$

Values of $F(K) = e^{-K^2}\int_0^K e^{z^2}\, dz$ have been tabulated (Miller & Gordon 1931) in the range $K = 0$ to $K = 12$. To use these tables, (26) can be expressed as

$$T_1 = 2[F(K) - \exp[-K^2 + (C/a)^2]\, F(C/a)]. \tag{27}$$

Since the plastic-elastic boundary moves with velocity C, $CT = L_1 - X$, which, in non-dimensional form, becomes

$$\frac{C}{a} T_1 = \frac{L_1}{L} - \frac{X}{L}, \tag{28}$$

while (24) becomes

$$\log_e (L/X) = K^2 - (C/a)^2. \tag{29}$$

Equations (27), (28) and (29), together with

$$K = U/a + C/a, \tag{30}$$

are sufficient to determine C/a, U/a, K and T, when X/L and L_1/L are known.

To estimate the error arising from the use of equation (22) instead of the more accurate equations (27) to (30) in any particular case where L_1 and X are measured, it is necessary to solve transcendental equations. This laborious calculation was carried out by Mr Whiffin in the case of several of the experimentally determined values of L_1 and X. It was found that, in all cases, equation (22) underestimated the yield stress as compared with the equations (27) to (30). Subsequently, the corrections were calculated systematically in the form of a correcting factor S/S_1, which, when applied to S_1, as calculated from (22), gives the value which would have been obtained for S if (27) to (30) had been solved.

Combining equations (27), (28) and (29)

$$\frac{L_1}{L} = 2\frac{C}{a}F(K) - \left[2\frac{C}{a}F(C/a) - 1\right]\frac{X}{L}. \tag{31}$$

Since

$$\frac{2S}{\rho U^2} = \frac{a^2}{U^2} = \frac{1}{(K-C/a)^2}$$

the correcting factor is

$$\frac{S}{S_1} = \frac{L-L_1}{L-X}\frac{\log_e (L/X)}{(K-C/a)^2}. \tag{32}$$

In performing the calculation, a value of X/L was taken, and a series of values of C/a covering the range from 0 to infinity were used to calculate the corresponding values of L_1/L and S/S_1. These were plotted on a diagram with ordinates S/S_1 and abscissae L_1/L. A series of curves were thus obtained, each corresponding with a given value of X/L. The values of L_1/L corresponding with definite values of S/S_1, namely, 1·0, 1·05, 1·10, ..., 2·0, were taken from these curves. In this way, and with the help of scattered values calculated by Mr Whiffin, the diagram of figure 4 was constructed. Here the ordinates are X/L and the abscissae L_1/L. The curves on the diagram are contours of equal correction factors. It will be seen that the curvature of these contours is very slight except near the axis $X/L = 0$.

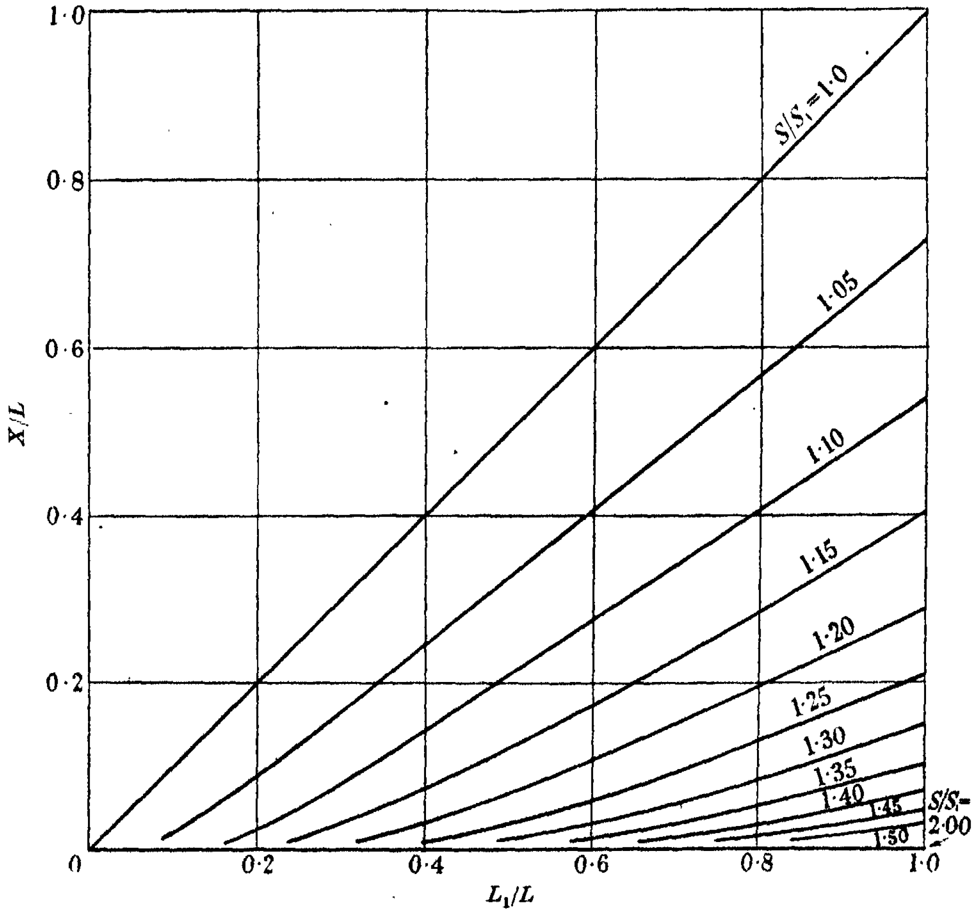


FIGURE 4. Contours of ratio of yield value S computed by exact formula to value S_1 obtained by approximate formula.

It will be noticed that the contours cut the line $L_1/L = 1·0$ without any apparent singularity. The limit points on $L_1/L = 1·0$ were calculated using the asymptotic form of $F(x)$ as $x \to \infty$, namely

$$\lim_{x \to \infty} F(x) = \frac{1}{2x} + \frac{1}{4x^3} + \dots \tag{33}$$

In this way it is found that, when C/a and K are infinite,

$$L_1/L = 1{\cdot}0, \tag{34}$$

and

$$\frac{S}{S_1} = 2\left[\frac{1}{1-X/L} - \frac{1}{\log_e(L/X)}\right]. \tag{35}$$

The limiting points on the line $L_1/L = 1{\cdot}0$ in figure 4 were calculated from equation (35).

EXPERIMENTS WITH PARAFFIN WAX

The method described above was first applied to find the dynamic yield stress of paraffin wax. Transparent cylinders of this material were cut from cast blocks, and these were projected by means of a catapult, capable of giving them a speed of 125 ft./sec., at a heavy anvil hung as a ballistic pendulum. The cylinders were 1·75 cm. long. After the impact they were found to be shorter, but they remained coherent. Paraffin wax has the property of remaining transparent under compressive stressing until a sudden collapse occurs. The material which has yielded is full of small cracks, which give it an opaque white appearance. The cylinders which had been projected were found to be opaque at the impact end, but they remained transparent at the rear end. The yield point was taken to correspond with the boundary between the transparent and opaque portions. The length of the transparent portion was taken as X. Static tests were also made by compressing paraffin wax cylinders between polished plates in a parallel-motion compression machine, and it was found that sudden and catastrophic breakdown occurred at a certain load, the wax remaining transparent up to the instant of breakdown.

Some results are given in table 3. It will be seen that the dynamic yield stress varied from 840 to 930 lb./sq.in., while the static yield stress was only 485 lb./sq.in. The ratio (dynamic yield stress)/(static yield stress) was therefore about 1·8.

TABLE 3. EXPERIMENTS WITH CYLINDERS OF PARAFFIN WAX

L	U	L_1	X	S dynamic yield stress from eqn. (22)	S static yield stress (mean of 6 observations)
(cm.)	(ft./sec.)	(cm.)	(cm.)	(lb./sq.in.)	(lb./sq.in.)
1·774	126	1·635	0·95	854	—
1·757	128	1·625	0·95	930	485
1·779	132	1·625	0·95	840	—

INTERPRETATION OF RESULTS

The measurements made by Mr Whiffin show that his specimens maintained the stresses, which are here called dynamic yield stresses, instantaneously without suffering strain greater than 0·2 %. When considering the results of mechanical tests at comparatively low speeds, in which the elastic limit is passed and plastic strain occurs, the rate of strain is definable in terms of the experimentally measured quantities. In fact, if l is the length of the specimen at time t, and L its original length, one definition of rate of strain is simply $1/L\,.\,dl/dt$. In experiments of the type here considered, it is possible to define the mean rate of strain of the plastically distorted portion of the projectile. If the approximation represented by (22) is adopted, the mean rate of strain is $U/2(L-X)$, because the reduction in length is $(L-L_1)$, and this is entirely confined to the material whose initial length was $(L-X)$. The total strain of the portion which has yielded is therefore $(L-L_1)/(L-X)$. Since the deceleration is assumed uniform, its duration is $2(L-L_1)/U$, so that the mean rate of strain is

$$\frac{(L-L_1)}{(L-X)}\frac{\frac{1}{2}U}{(L-L_1)}=\frac{\frac{1}{2}U}{(L-X)}. \tag{36}$$

Values of this mean rate of strain are given in Mr Whiffin's paper because they are the only rates of strain which can be deduced from his measurements. It must, however, be remembered that, at all stages of the impact, the analysis refers only to the part of the projectile which has not yet suffered plastic compression, so that the connexion between the rate of strain just defined and the yield stress can only be an indirect one.

It seems impossible to derive further information about the physical factors which determine the dynamic yield stress without making more complete measurements of the successive states of the projectile during impact. It is, however, worth noticing that, if the deceleration of the rear portion of the projectile is continuous, as is contemplated in equations (5), (6) and (7), the maximum stress S at a distance X from the rear end is only attained instantaneously at the end of the impact. At the beginning of the impact, the stress in this part of the projectile is XS/L. If the stress at distance X rises uniformly from the value XS/L at the beginning to S at the end, the stress at time t would be

$$S\left[\frac{X}{L}+\frac{t}{T}\frac{(L-X)}{L}\right].$$

The duration of stress greater than, say, $S(1-y)$, where y is small, is $L/(L-X)\,.\,Ty$. To a rough approximation, therefore, one may expect that, at the place where the yield point is found, the stress has exceeded, say, 99 % of S for a duration of the order of $(0{\cdot}01)\ L(L-X)\,.\,T$, and, according to the simple theory of equation (22), $T=2(L-L_1)/U$. In one group of Mr Whiffin's experiments, recorded in his table 2, the estimated value of T lies between $5{\cdot}2$ and $7{\cdot}0\times10^{-5}$ sec., so that there is very little variation in the duration of the impact, although there is great variation in U and $(L-L_1)$. It seems possible that the constancy of S, which is found for varying velocities of the projectile, is due to the constancy of the duration of impact.

REFERENCE

Miller, W. L. & Gordon, A. R. 1931 *J. Phys. Chem.* **35**, 2878.

The use of flat-ended projectiles for determining dynamic yield stress

II. Tests on various metallic materials

By A. C. Whiffin

The Road Research Laboratory, Department of Scientific and Industrial Research

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 24 *July* 1947)

A description is given of the experimental technique devised to apply the method outlined theoretically in part I to the measurement of the dynamic compressive yield strength of various steels, duralumin, copper, lead, iron and silver.

A polished piece of armour steel was employed as a target, and cylindrical specimens were fired at it at various measured velocities from Service weapons. The distance between the weapon and target was made short to ensure normal impact, and apparatus was devised for the precise measurement of striking velocity over this short range. The dynamic compressive yield strength was computed from the density of the specimen, the striking velocity, and from measurements of the dimensions of the test piece before and after test.

Details are given of the accuracy of the various measurements, and of their effect on the values of yield strength. The method was found to be inaccurate at low and high velocities. For instance, with mild steel, satisfactory results were only obtainable within the range 400 to 2500 ft./sec. The range of velocities within which satisfactory results could be obtained varied with the quality of the material tested, soft metals giving results within a much lower range than that necessary for harder materials. Because of its failure at low velocities, the method could not be employed to bridge the gap between static and dynamic tests. The rate of strain employed in the dynamic tests could not be measured, but was estimated to be of the order of 10,000 in./in./sec.

With the materials tested little change of dynamic strength occurred within the range of striking velocities employed, probably because the rate of strain did not vary to any great extent with the striking velocity. Within the range of weapons available, that is, from a 0·303 in. rifle up to a 13 pdr. gun (calibre 3·12 in.), little change of dynamic strength occurred with alteration of the initial dimensions of the specimens, probably because the corresponding change of rate of strain was not large.

In general, the dynamic compressive yield strength S was greater than the static strength Y represented by the compressive stress giving 0·2 % permanent strain. For steels of various types, regardless of chemical composition and heat treatment, there was a relation between S/Y and the static strength Y, the ratio decreasing from approximately 3 when Y was 20 tons/sq.in. to 1 when Y was 120 tons/sq.in. A similar relation occurred with duralumin, S/Y varying from 2·5 at $Y = 8$ tons/sq.in. to 1·4 at $Y = 25$ tons/sq.in.

Dynamic compressive yield values were obtained for soft materials such as pure lead, copper and Armco iron, which, under static conditions, gave no definite yield values.

A plot of the unstrained length of the specimen X, expressed as X/L (where L = initial overall length), versus the final overall length L_1, expressed as L_1/L, was made for the various materials. Any specified value of X/L was associated with greater values of L_1/L for the more ductile materials, such as copper and lead, than for the brittle materials, such as armour plate and duralumin.

Introduction

During the period 1939–40, the Road Research Laboratory of the Department of Scientific and Industrial Research was engaged upon experimental work for the Research and Experiments Department of the Ministry of Home Security, and it became necessary to determine the strength of various materials at the high rates of strain associated with the impact and penetration of projectiles. The rates of strain obtainable with impact machines, up to approximately 900 in./in./sec. (Brown & Vincent 1941) were considered to be too low for the purpose, and hence the method outlined in part I by Sir Geoffrey Taylor, F.R.S., was adopted.

Preliminary tests made with mild steel showed that the dynamic yield strength was considerably above the static value. The results were brought to the notice of the Service Departments, where it was appreciated that the technique afforded a means of testing steels of the type employed in constructing shells, bombs, and armour plate, the conditions of test being at least similar to those obtaining in practice. Consequently, much of the work described in this report was conducted for the Ministry of Supply, and the bulk of it was sponsored by the Armour-Piercing Projectile Co-ordinating Sub-Committee of the Ordnance Board under the Chairmanship of Dr C. Sykes, F.R.S., at that time Superintendent of the Metallurgy Division of the National Physical Laboratory.

Tests were also made on other materials such as duralumin, copper, lead, and iron, while tests on duralumin and silver were made for the Metallurgy Division of the National Physical Laboratory, and their metallographic examination is described in part III.

EXPERIMENTAL PROCEDURE

In many of the experiments the steel specimens were 1·00 in. long by 0·280 in. diameter, electroplated with copper to increase the diameter to 0·312 in. for firing from a 0·303 in. Service rifle. The projectiles were fired to strike perpendicularly upon the ground and polished surface of a piece of cemented steel armour plate, 12 in. square by 4 in. thick. The specimens were usually 'mushroomed' by the impact, and the profiles of typical specimens are shown in figure 1.

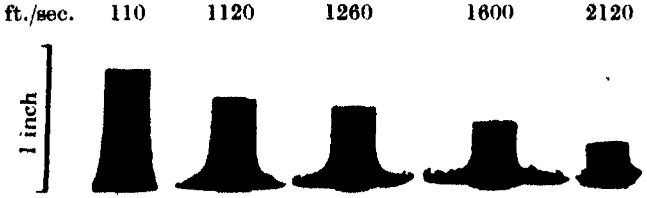


FIGURE 1. Steel cylinders after test.

In much of the work described in this paper, the approximate formula (equation (22) of part I) has been used to compute the dynamic compressive yield strength from the experimental data. This formula, converted to practical units, is

$$S = \frac{1{\cdot}018 \times 10^{-5} \times U^2}{\log_{10}(L/X)} \frac{L-X}{L-L_1+d}, \tag{1}$$

where S = dynamic compressive yield strength in tons/sq.in.,

U = striking velocity in ft./sec.,

L = initial overall length of the test cylinder—usually measured in inches,

L_1 = overall length of the test cylinder after impact—measured in similar units to L,

X = unstrained length of the specimen in inches—taken as the distance from the rear end of the specimen to the point where the diameter first increases by 0·001 in.,

d = depth of dent in the target plate, in inches—frequently negligible.

In equation (1), the constant $1{\cdot}018 \times 10^{-5}$ refers to steel having a density of 7·80 g./ml. When other materials are tested, the constant must be adjusted in proportion to the densities.

In the more exact analysis given in part I (see equations (27) to (30)), the following equations have to be solved:

$$\int_{C/a}^{K} e^{z^2}\,dz = \frac{L_1/L - X/L}{2C/a}\,e^{K^2}, \tag{2}$$

$$\log_e (L/X) = K^2 - \frac{C^2}{a^2}. \tag{3}$$

K, C, and a are constants, given by

$$2S/\rho = a^2 \quad \text{and} \quad K = \frac{U+C}{a}, \tag{4}$$

where ρ = density, and, when T = duration of impact, and $T_1 = aT/L$,

$$\frac{CT_1}{a} = \frac{L_1 - X}{L}. \tag{5}$$

The constant C represents the velocity of the plastic-elastic boundary in the specimen, relative to the front face of the target plate. The method of solution involves finding a value of K to satisfy equation (2), after which C/a, and then C and a are computed, and hence S is determined.

Both methods of solution depend on the striking velocity and on the dimensions of the test cylinder before and after impact. The overall lengths of the cylindrical specimens before and after test were measured by a micrometer to an accuracy of 0·001 in. After test, the copper coating was removed without damaging the steel, and measurements were made of the unstrained length X, to an accuracy of 0·005 in., by determining with dividers the distance from the rear end at which the diameter of 0·280 in. increased by 0·001 in. The unstrained length determined in this way was associated, therefore, with a permanent strain of similar order to that employed by the Engineering Division of the National Physical Laboratory in determining the stress giving 0·2 % permanent strain in static compression, this quantity being determined for most of the materials mentioned in the present paper. The unstrained length was determined at several points around the specimen, and the average value was used when substituting in the various formulae. It is shown in part III that metallographic examination confirmed the values of unstrained length determined by this technique.

The approximate formula (1) for determining dynamic strength involves the square of the striking velocity, while it will be shown later that the strength value obtained by the more exact analysis is almost equally dependent upon the striking velocity. Consequently, it was necessary to devise a method of determining the striking velocity as accurately as possible. In order to minimize yaw, most of the tests were made with a distance of only 5 to 10 ft. between the weapon and the target plate, and none of the Service chronographs available at that time, such as the Boulengé chronograph, was suitable for measuring velocities over such a short range. The method employed at the Road Research Laboratory was based on a circuit originally suggested by a member of the staff, Mr K. L. C. Freeborn, and consisted in timing the interval between the rupture of two wire screens by the passage of the projectile.

The wire screens were usually placed either 2 or 5 ft. apart, depending on whether the velocity to be used was below, or above, approximately 1000 ft./sec. The screens were wound in the form of a flat grid, with paper attached to both the front and rear faces by photographic dry-mounting tissue to prevent the wires moving laterally out of the path of the projectile without being ruptured. The spacing between the wires was less than the calibre of the projectile, being of the order of $\frac{1}{8}$ in. for the 0·303 in. projectile.

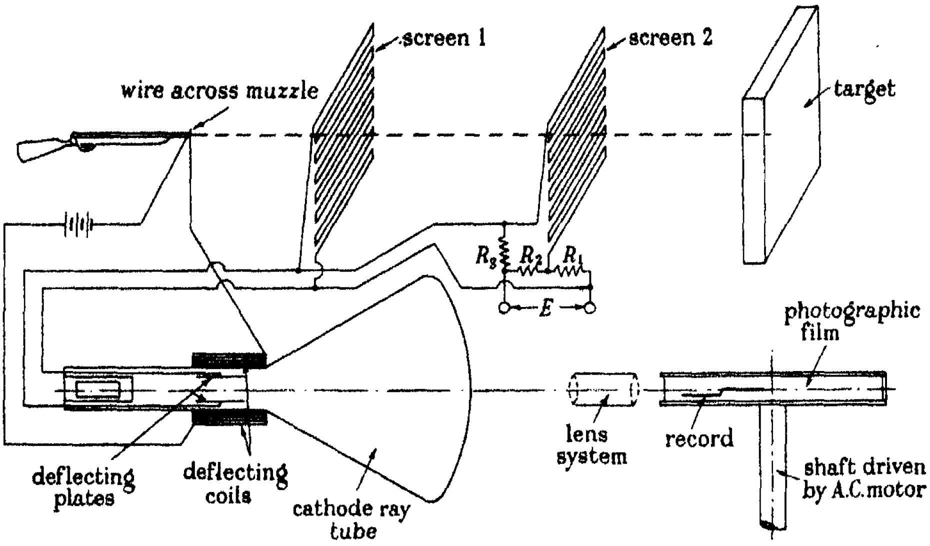


FIGURE 2. Details of method of measuring velocity of projectiles.

The recording circuit is shown in figure 2. It consisted essentially of a potential-divider connected across the output of a mains rectifier producing approximately 150 V, and connected to one pair of the deflecting plates of a cathode-ray tube. The movement of the spot on the screen of the tube was recorded photographically by a rotating-drum camera, driven by a lightly loaded induction motor, which, consequently, maintained a constant speed. In order to prevent fogging of the record due to the spot being in the centre of the screen when the camera shutter was opened, the spot was initially deflected off the screen by means of an electromagnetic coil wound around the neck of the tube, and connected in series with a battery and a wire placed across the muzzle of the weapon (see figure 2). When the weapon was fired, the muzzle wire was broken, either by blast or by the projectile, and the cathode-ray spot moved to its zero position in the centre of the tube screen. Rupture of the first wire screen applied a potential of $150R_1/[R_1+R_2R_3/(R_2+R_3)]$ to the deflecting plates of the tube, and caused the spot to move a short distance perpendicular to the direction of travel of the photographic film. Rupture of the second wire screen applied the full 150 V to the plates and deflected the spot off the screen. A typical record is shown in figure 3.

The velocity of the projectile was determined from the distance between the screens, and the length of the record corresponding to the interval between the rupture of the two wire screens. The drum camera was calibrated at frequent intervals by applying the output of a valve-maintained 1000-cycle tuning fork to the deflecting plates of the cathode-ray tube, and the writing speed remained constant at 13·4 mm./msec.

FIGURE 3. Typical record from velocity recording unit.

Length of trace (between points 1 and 2) = 16·52 cm.
Writing speed of unit = 1·34 cm./msec.
Duration of trace = 12·28 msec.
Screen distance = 5 ft.
Velocity of projectile = 5 × 1000/12·28.
= 407 ft./sec.

To avoid the necessity of developing the photographic record after each shot, the movement of the spot due to the rupture of the first wire screen was varied by using a rheostat in place of R_1, and placing a reversing switch between the bridge circuit and the cathode-ray tube to permit deflexions in both the upward and downward directions. It was thus possible to obtain as many as six to ten distinguishable records on one film.

The longer records, corresponding with the lower projectile velocities, were measured by a rule and dividers, while short records were measured by a travelling microscope, the overall accuracy being, in general, of the order of 1 in 500. The distance between the wire screens could be determined to an accuracy of $\frac{1}{4}$ in. when the screen distance was 5 ft., and to $\frac{1}{10}$ in. when the screen distance was 2 ft. It is considered, therefore, that the overall accuracy of the velocity measurements was approximately ± 5 ft./sec. within the velocity range employed, which was from 250 to 2750 ft./sec.

Allowing for the errors in velocity measurement and in determining the various dimensions of the test specimens, it is considered that the value of the dynamic compressive yield strength obtained in a single experiment, and computed by the approximate formula (see equation (1) of this paper), was accurate to approximately ± 5 % at the centre of the velocity range. In order to compare the effect of experimental errors on the dynamic strengths computed from the more exact analysis and from the approximate formula, calculations have been made of the effect of a change of 0·01 in X/L and L_1/L, and of 10 ft./sec. in the striking velocity, each change being examined independently. Such calculations have been made for mild steel specimens tested at 400, 1600 and 2500 ft./sec., and the results are summarized in table 1.

TABLE 1. EFFECT OF EXPERIMENTAL ERRORS

			results obtained by							
			approx. equation		more exact method of analysis					
striking velocity (ft./sec.)	X/L	L_1/L	dynamic strength S_1 (tons/sq.in.)	diff. due to error (tons/sq.in.)	K	a (ft./sec.)	C (ft./sec.)	duration of impact $T\times10^5$ (sec.)	dynamic strength S (tons/sq.in.)	diff. due to error (tons/sq.in.)
400	0·45	0·94	43·04	—	1·5700	1433	1850	2·21	48·13	—
400	0·44	0·94	42·62	−0·42	1·6050	1428	1891	2·20	47·75	−0·38
400	0·45	0·95	51·65	8·61	1·6968	1572	2268	1·84	57·93	9·80
410	0·45	0·94	45·22	2·18	1·5700	1469	1896	2·15	50·57	2·44
1600	0·14	0·41	44·47	—	1·4420	1447	487	4·62	49·09	—
1600	0·13	0·41	44·36	−0·11	1·4720	1433	510	4·58	48·15	−0·94
1600	0·14	0·42	45·24	0·77	1·4455	1462	514	4·54	50·10	1·01
1610	0·14	0·41	45·03	0·56	1·4420	1456	490	4·59	49·70	0·61
2500	0·03	0·13	46·57	—	1·8790	1450	225	3·71	49·27	—
2500	0·02	0·13	42·17	−4·40	1·98632	1386	253	3·62	45·03	−4·24
2500	0·03	0·14	47·11	0·54	1·88033	1462	249	3·68	50·10	0·83
2510	0·03	0·13	46·94	0·37	1·8790	1456	226	3·70	49·67	0·40

The changes in dynamic strength given in table 1 for the various changes in X/L and L_1/L are not directly related to the striking velocities shown, but to the initial values of X/L and L_1/L. Other metals will give similar values of X/L and L_1/L at velocities different from those shown, depending on the strength of the material tested.

The principal conclusion to be drawn from table 1 is that the effects of experimental errors on the results of the more exact analysis are of a similar order to those of the approximate method. The effect of errors in measuring the dimensions of the test-pieces is to produce a decrease of accuracy towards both ends of the velocity range. At high velocities the unstrained length X approaches zero, and small variations in X have a large effect on the term $\log(L/X)$. This is shown in table 1, where a change of 0·01 in X produced a change of 10 % in the dynamic strength obtained at 2500 ft./sec., compared with a change of 0·2 to 2·0 % at 1600 ft./sec., and a change of 1 % at 400 ft./sec. At low velocities, the final overall length L_1 differs little from the initial overall length L, making the term $(L-L_1+d)$, used in the approximate formula, very susceptible to small changes in L_1; similar fluctuations occur in the results of the exact analysis. In table 1, for instance, a change of 0·01 in L_1 caused a change in dynamic strength of 20 % at 400 ft./sec., compared with a change of 1·7 to 2·0 % at 1600 ft./sec., and 1·1 to 1·7 % at 2500 ft./sec. The experiments become valueless, therefore, at high and at low velocities, and the range of velocities within which they have value will depend upon the quality of the material tested. A weak, soft material will give satisfactory results within a lower velocity range than a strong material such as armour steel.

Table 1 shows that the effect of variation of striking velocity diminishes with increase of velocity, as might be expected.

It will have been noticed from table 1 that the exact analysis gave greater values of dynamic strength than the approximate formulae, the difference being 11·8 % at 400 ft./sec., 10·3 % at 1600 ft./sec., and 5·8 % at 2500 ft./sec. Differences of this magnitude are outside the range of experimental error expected in this work. It is clear from the formulae employed that the percentage increase in yield strength, due to employing the exact method instead of the approximate formula, depends solely on the values of X/L and L_1/L, and is independent of the striking velocity, the density of the projectile, and the initial length of the specimen, and figure 4 of part I shows contours of constant percentage change plotted against X/L and L_1/L. By using this diagram it was possible to adjust the values of dynamic strength computed by the approximate formula to correspond with the exact analysis without the labour of performing the exact analysis for each individual experiment, and this technique has been used in all the experiments described in the present paper.

ESTIMATION OF RATE OF STRAIN

In the experiments described in this paper it has not been possible to obtain measurements of the rate of strain. Several attempts were made to determine the rate by photographic methods, but the relatively small dimensions of the specimens made it difficult to obtain results by any of the high-speed photographic methods available at the laboratory.

If it may be assumed that the deceleration of the projectile is uniform, then the rate of strain may be estimated by using equation (36) of part I, which becomes

$$\text{Rate of strain} = \frac{6U}{L-X}\ \text{in./in./sec.,} \tag{6}$$

when the striking velocity, U, is in ft./sec. and L and X are measured in inches.

For the approximate solution the duration of impact, T, is given by

$$T = \frac{L-L_1}{6U}\,\text{sec.,} \tag{7}$$

while for the exact solution the duration may be obtained from equation (5).

VARIATION OF DYNAMIC COMPRESSIVE YIELD STRENGTH WITH STRIKING VELOCITY

In all the tests made with any particular size of specimen and type of material, although the dynamic compressive yield strength was usually above the static 0·2 % compressive proof stress, it did not usually vary significantly with striking velocity within the range investigated. For instance, a series of tests was made within the velocity range 1410 to 2750 ft./sec., using mild steel cylinders having a diamond pyramid hardness of 120, the specimens being initially 0·280 in. in diameter by 1·00 in. long. The dynamic compressive yield strength remained approximately constant at 47 tons/sq.in. throughout the range of velocities employed, and was greatly in excess of the value of 18 tons/sq.in. obtained by the National Physical Laboratory for the 0·2 % compressive proof stress. Table 2 gives typical values of the specimen dimensions and of the dynamic compressive yield strengths obtained in these tests, together with estimated values of the rates of strain and durations of impact, calculated by equations (6) and (7).

TABLE 2. RESULTS OF DYNAMIC TESTS ON MILD STEEL (D.P.H. = 120)

striking velocity U (ft./sec.)	overall length of specimen after test L_1 (in.)	unstrained length X (in.)	duration of impact (approx. formula) T (sec. × 10^5)	estimated rate of strain (in./in./sec.)	dynamic compressive yield strength by approx. formula S_1 (tons/sq.in.)	dynamic compressive yield strength by exact method S (tons/sq.in.)
1410	0·429	0·26	6·7	11,400	44·8	47
1480	0·400	0·22	6·8	11,400	44·1	47
1570	0·343	0·20	7·0	11,800	43·7	46
1570	0·360	0·19	6·8	11,600	44·2	47
1730	0·303	0·16	6·7	12,400	46·2	49
1850	0·263	0·13	6·6	12,800	46·4	49
2045	0·212	0·09	6·4	13,500	47·0	50
2150	0·181	0·06	6·4	13,700	44·2	47
2350	0·146	0·03	6·1	14,500	41·9	44
2750	0·141	0·01	5·2	16,700	44·4	47

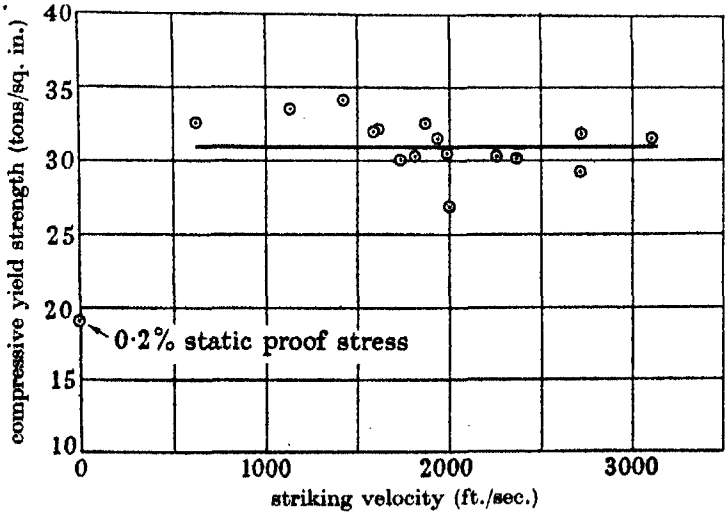


FIGURE 4. Dynamic-compressive yield strength of duralumin projectiles.

If it be assumed that the static test to determine the compressive stress corresponding with 0·2 % permanent strain takes approximately 15 min., the rate of strain is approximately 2×10^{-6} in./in./sec. It is not surprising, therefore, that, since the increase of the rate of strain from 2×10^{-6} to 13,000 in./in./sec. raised the yield from 18 to 47 tons/sq.in., variation of the striking velocity from 1410 to

2750 ft./sec., which produced an increase in the rate of strain from 11,400 to 16,700 in./in./sec., had no significant effect on the dynamic compressive yield strength.

Figure 4 shows a typical set of results obtained with duralumin cylinders, again showing that the dynamic strength was independent of the striking velocity within the range employed.

EFFECT OF DIMENSIONS OF SPECIMENS

It has been mentioned already that the range of velocities which can be used to give satisfactory results depends on the quality of the material tested, and, therefore, with soft materials this velocity range will be well below that for the various types of steel. Consequently, the range of velocities (1300 to 2750 ft./sec.) usually obtainable with the 0·303 in. rifle was not satisfactory for tests on soft materials. Attempts were made to fire the rifle at velocities below 1300 ft./sec., but they frequently resulted in the projectile remaining fixed in the barrel.

A solution of the problem was to employ other weapons to obtain the lower velocity range required, but since, for most purposes, this involved using a 0·38 in. revolver with steel cylinders 0·350 in. in diameter by 0·500 in. long, it was first necessary to determine the effect on the results of varying the dimensions of the specimens, particularly since it was likely that both the rifle and the revolver would have to be used to obtain a suitable velocity range for some of the materials investigated.

Inspection of the formulae derived by Sir Geoffrey Taylor showed that the dynamic strength should be independent of the calibre of the cylindrical specimen, and should also be independent of the length, provided that the unstrained length X and the overall length L_1 were proportional to the initial length L, and that the rate of strain was not greatly affected by varying the initial dimensions of the test cylinder. Figure 5 shows the similarity between two cylinders after test at 575 ft./sec., one being initially 3·12 in. in diameter by 4·00 in. long, while the other was 0·35 in. in diameter by 0·50 in. long.

The effect of specimen length was investigated in a series of tests on cylinders of hard mild steel (diamond pyramid hardness = 170). The cylinders were 0·500 in. in diameter by 0·80 to 1·60 in. long, varying by intervals of 0·20 in. The specimens were electroplated in the usual manner and fired from a 0·55 in. anti-tank rifle. The results are summarized in table 3.

Table 3 shows that increase of the initial length of the specimen from 0·80 to 1·60 in. made no significant difference to the dynamic compressive yield strength, the slight change recorded, from 54 to 57 tons/sq.in., being less than the experimental scatter. The rate of strain decreased with increase of the length of the test cylinder but remained high, and the range covered was such that this variation in rate of strain would not be expected to affect the dynamic strength.

In connexion with other work it became necessary to determine the relation between muzzle velocity and weight of propellant for a 13 pdr. gun. The proof shot employed consisted of a mild steel cylinder 3·12 in. in diameter by 4·00 in. long, which was fired against a piece of armour plate 2 in. thick by 3 ft. square. The velocities employed lay between 575 and 1200 ft./sec., and the 'mushroomed' cylinders were sufficiently similar to those obtained in other tests to justify making cylinders from the material to fit other weapons, so as to determine the effect of specimen size. One of the large cylinders after test is shown in figure 5; the increase

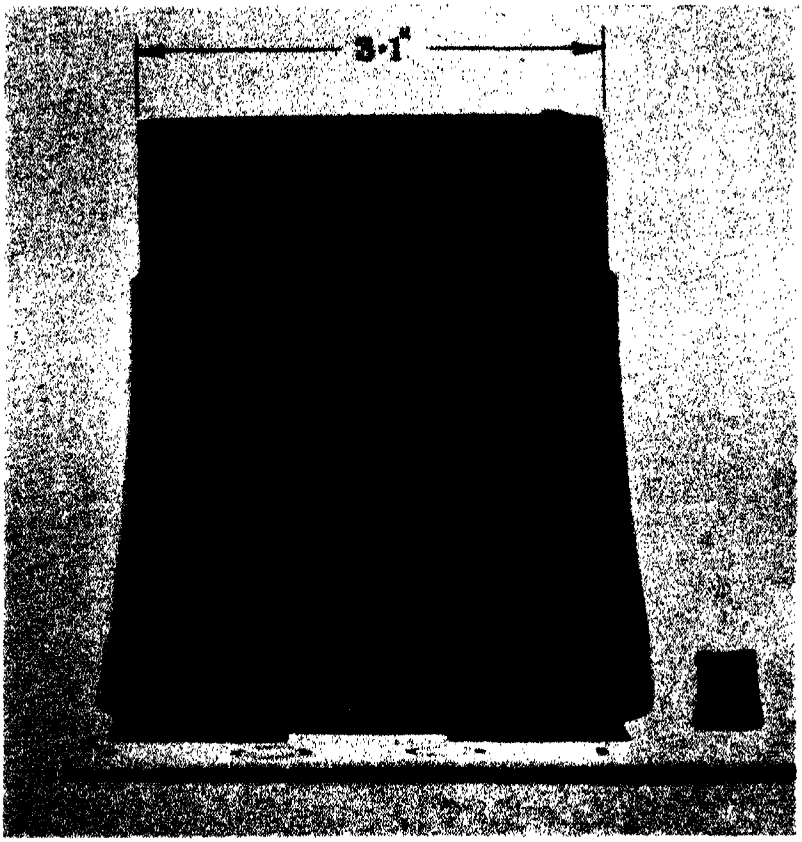


FIGURE 5. Large and small mild steel cylinders after test at 575 ft./sec.

TABLE 3. EFFECT OF LENGTH OF SPECIMEN: TESTS WITH CYLINDERS OF 0·50 IN. DIAMETER (MILD STEEL; D.P.H. = 170)

			average values of		
				dynamic compressive yield strength	
initial length of specimen (in.)	range of striking velocities employed (ft./sec.)	range of rates of strain obtained (in./in./sec.)	rate of strain (in./in./sec.)	by approx. formula (tons/sq.in.)	by exact analysis (tons/sq.in.)
0·80	1960–2235	16,200–17,800	17,200	48	54
1·00	1740–2310	12,400–14,700	13,200	49	55
1·20	1540–2160	10,000–11,600	10,600	51	55
1·40	1540–1650	8,300– 8,900	8,700	51	55
1·60	1470–1990	7,600– 8,600	8,000	53	57

in diameter at about a quarter of the way along the specimen was made to fit a cartridge case, and was not caused by the dynamic experiment. The steel had a diamond hardness of 120, and the dimensions of the various cylinders are given in table 4 together with the results.

TABLE 4. EFFECT OF SPECIMEN DIMENSIONS: TESTS ON MILD STEEL

dimensions of test cylinder		range of striking velocities employed (ft./sec.)	range of rates of strain obtained (in./in./sec.)	average values of		
					dynamic compressive yield strength	
diam. (in.)	length (in.)			rate of strain (in./in./sec.)	by approx. formula (tons/sq.in.)	by exact analysis (tons/sq.in.)
0·280	1·000	1440–2750	10,800–16,700	12,800	45	47
0·350	0·500	380–1125	8,300–15,800	13,700	45	52
0·500	1·000	1100–2140	8,700–13,800	11,200	47	54
3·120	4·000	575–1200	1,150– 1,810	1,430	39	48

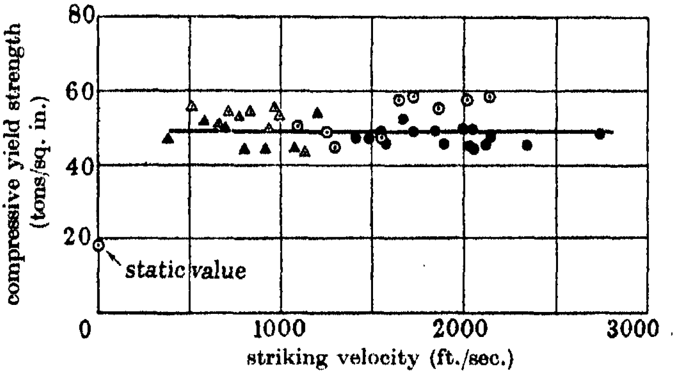


FIGURE 6. Effect of specimen dimensions on the dynamic compressive yield strength of cylindrical projectiles of mild steel.

	dimensions of specimens (in.)	
	diam.	length
●	0·280	1·000
⊙	0·500	1·000
▲	3·120	4·000
△	0·350	0·500

The results of the tests are shown in figure 6, and indicate that the specimen dimensions had no marked effect on the values of dynamic strength. The scatter of the experimental results was similar in magnitude to that which occasionally occurs in this work, and was greater than that expected from the possible errors in measurement.

COMPARISON OF APPROXIMATE AND EXACT METHODS OF ANALYSIS FOR MILD STEEL TESTED WITHIN THE VELOCITY RANGE 380 TO 2750 FT./SEC.

The experiments described in the previous section showed that the specimen dimensions had little effect on the dynamic strength. Accordingly values of X/L and L_1/L were plotted against the striking velocity for these experiments, and these quantities were found to be independent of the initial dimensions of the specimens. Values read off these curves are shown in figure 7 (*a*), together with values of $(L_1-X)/L$. Since these experiments covered a greater velocity range than any of the others described in the present paper, the results have been analyzed by both the approximate and the exact methods to show the difference which arises, and also to show the variation of K, C, and a with striking velocity. The results of the analysis are summarized in table 5.

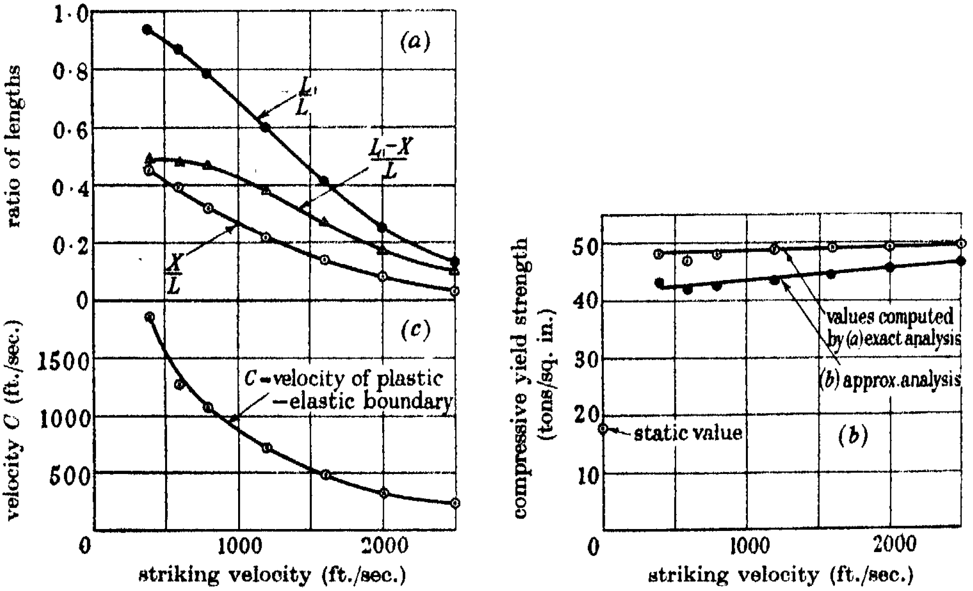


FIGURE 7. Results of tests on mild steel cylinders.

The results of the two methods of analysis are also shown graphically in figure 7, and indicate that the differences between the two sets of values of dynamic strength decrease with increase of striking velocity. However, this result is only applicable to the particular values of X/L and L_1/L obtained in this set of experiments. It has been shown in figure 4 of part I that the difference between the approximate and exact solutions depends solely on the actual values of X/L and L_1/L employed.

One interesting feature of table 5 is the near constancy of the term a. This was to be expected since $2S/\rho = a^2$, and the dynamic strength S has been found in these experiments to be independent of the striking velocity.

Figure 7 (*c*) shows the variation with striking velocity of the velocity C of the plastic-elastic boundary with reference to the target plate. C appears to decrease with increase of striking velocity, and to approach zero asymptotically at high striking velocities. This appears reasonable because, at very high striking velocities, the projectile probably behaves as a liquid jet, and becomes spread over the face of the target, so that the plastic-elastic boundary cannot move through the projectile away from the target face. At low striking velocities it appears possible that the velocity of the boundary will approach that of elastic waves in the specimen.

TABLE 5. COMPARISON OF RESULTS OF APPROXIMATE AND EXACT METHODS OF ANALYSIS: MILD STEEL TESTED OVER RANGE 380 TO 2750 FT./SEC.

(0·2 % compressive proof stress = 18 tons/sq.in.)

			results obtained by							
			approx. method		more exact method of analysis					increase in dynamic strength due to using exact method (%)
striking velocity (ft./sec.)	X/L	L_1/L	rate of strain (in./in./sec.)	dynamic strength S (approx.) (tons/sq.in.)	time* of impact T (sec. $\times 10^5$)	K	a (ft./sec.)	C (ft./sec.)	dynamic strength S (exact) (tons/sq.in.)	
400	0·45	0·94	4,400	43·1	2·2	1·5700	1433	1850	48·1	11·8
600	0·39	0·87	5,900	42·0	3·1	1·3246	1419	1279	47·1	12·2
800	0·32	0·79	7,100	42·6	3·7	1·3000	1434	1064	48·2	13·0
1200	0·22	0·60	9,200	43·4	4·4	1·3263	1443	714	48·8	12·3
1600	0·14	0·41	11,200	44·5	4·6	1·4420	1447	487	49·1	10·3
2000	0·08	0·25	13,000	45·5	4·4	1·6048	1447	322	49·1	7·8
2500	0·03	0·13	15,500	46·6	3·7	1·8790	1450	225	49·3	5·8

* *The rates of strain and times of impact given in table 5 are based on an initial length of specimen of 1 in.*

Figure 7 (*b*) shows that, even at the lowest striking velocity employed, there was no tendency for the dynamic strength to fall towards the static value, probably because, as has been mentioned already, the lowest striking velocity employed was still associated with a high rate of strain compared with that employed in static experiments.

TESTS ON VARIOUS TYPES OF STEEL—CORRELATION BETWEEN DYNAMIC AND STATIC STRENGTHS

Table 6 gives details of the various types of steel which were tested, together with the hardness value, if measured, the 0·2 % proof stress obtained in static compression tests by the Engineering Division of the National Physical Laboratory, the average rate of strain attained in the dynamic tests, the average dynamic compressive strength and the ratio of dynamic to static strengths. The dynamic strengths were computed by the approximate method and then corrected to the exact formula by using the difference contour curve given in figure 4 of part I. The range of velocities employed was usually from 1400 to 2500 ft./sec., and average values of dynamic strength and rate of strain have been given in table 6 because, in general, the variation of dynamic strength with striking velocity was negligible.

Figure 8 (*a*) shows the ratio of dynamic to static strength plotted against the logarithm of the static strength, while figure 8 (*b*) shows the ratio plotted directly against the static strength. Reference to table 6 shows that the dynamic tests were

TABLE 6. SUMMARY OF RESULTS OF TESTS ON STEEL

reference number	details of material	hardness (B. = Brinell, D. = Diamond)	average estimated rate of strain in dynamic tests (in./in./sec.)	static strength (0·2 % proof stress in compression) Y (tons/sq.in.)	dynamic compressive yield strength S (tons/sq.in.)	S/Y
1	mild steel	121 (B.)	12,200	13	42	3·23
2*a*	do.	120 (D.)	13,200	18	49	2·72
2*b*	do.	120 (D.)	13,000	18	47	2·61
3	medium carbon steel (0·3 % C)	—	13,700	19	65	3·42
4	sample from base of shell	—	13,900	27	66	2·44
5	do.	—	14,400	27	70	2·59
6	armour plate: specimens cut parallel to face	250 (D.)	13,800	31½	70	2·22
7	plate as in number 6: specimens cut perpendicular to face	246 (D.)	15,200	33½	89	2·66
8	armour plate: specimens cut parallel to face	268 (D.)	14,600	32	78	2·44
9	plate as in number 8: specimens cut perpendicular to face	263 (D.)	14,200	33	77	2·33
10	sample from base of shell	—	15,200	33½	81	2·42
11	armour plate: specimens cut perpendicular to face	—	15,900	33	82	2·48
12	plate as in number 11: specimens cut parallel to face	—	14,000	37½	68	1·81
13	sample from base of shell	—	13,900	38	76	2·00
14	armour plate: specimens cut parallel to face	—	14,400	38	72	1·89
15	plate as in number 14: specimens cut perpendicular to face	—	14,000	39	90	2·31
16	sample from shoulder of shell	—	14,800	39	72	1·85
17	armour plate: specimens cut parallel to face	—	15,600	39	79	2·03
18	plate as in number 17: specimens cut perpendicular to face	—	15,600	41	84	2·05
19	armour plate: specimens cut parallel to face	—	14,200	44	71	1·61
20	plate as in number 19: specimens cut perpendicular to face	—	13,600	47	73	1·55
21	sample from shoulder of shell	—	16,000	52	98	1·88
22	do.	—	16,000	52	98	1·88
23	do.	—	16,200	53	103	1·94
24	nickel-chromium steel (3·55 % Ni, 0·47 % Cr)	—	15,300	53	85	1·60
25	sample from shoulder of shell	—	17,000	58	111	1·91
26	armour plate: specimens cut parallel to face (longitudinal)	—	15,300	59	92	1·56
27	plate as in number 26: specimens cut parallel to face (transverse)	—	16,400	60	111	1·85
28	plate as in number 26: specimens cut perpendicular to face	—	15,800	59	106	1·80
29	armour plate: specimens cut perpendicular to face	—	17,600	72	114	1·58
30	plate as in number 29: specimens cut parallel to face (longitudinal)	—	16,500	73	105	1·44
31	plate as in number 29: specimens cut parallel to face (transverse)	—	17,100	73	115	1·58
32	armour plate: specimens cut perpendicular to face	—	15,900	71	96	1·35
33	plate as in number 32: specimens cut parallel to face (longitudinal)	—	16,200	73	103	1·41
34	plate as in number 32: specimens cut parallel to face (transverse)	—	16,700	73	109	1·49
35	'Vibrac' steel	—	16,500	74	105	1·42
36	nickel-chromium steel	514 (D.)	22,000	120	133	1·11

made at an approximately constant rate of strain of the order of 15,000 in./in./sec., while the static tests were made at the rate of 1 or 2×10^{-8} in./in./sec. Figure 8 shows that increasing the rate of strain by approximately 10^{10} increased the yield strength

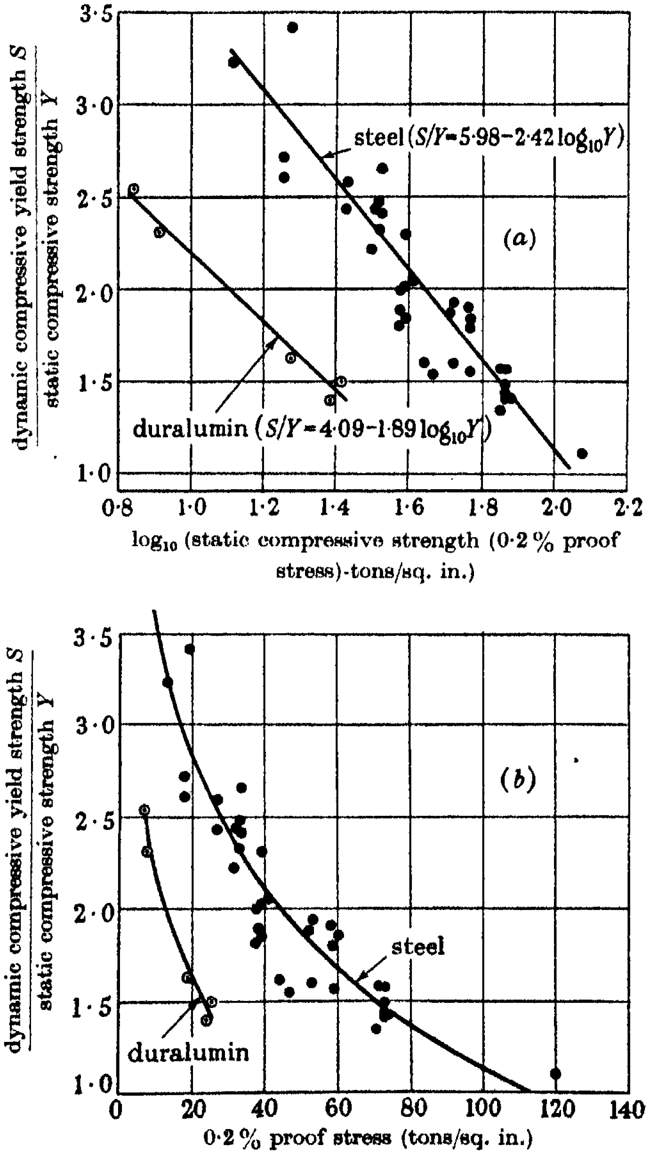


Figure 8. Variation of dynamic yield strength with static strength.

to approximately treble its static value at low static strengths, such as 20 tons/sq.in., while the increase of strength decreased with increase of static strength until there was no apparent increase at the highest static strength obtainable in the present experiments, namely, 120 tons/sq.in. The scatter shown in figure 8 is to be expected in view of the different types of steel employed, some of which contained alloying elements not present in the others, while some had been subjected to complicated heat treatments.

Figure 8 (*a*) indicates a linear relationship between the ratio of dynamic to static strengths and the logarithm of the static strength, the line on the diagram for steel being given by

$$S/Y = 5\cdot98 - 2\cdot42 \log_{10} Y, \tag{8}$$

where S = dynamic compressive yield strength in tons/sq.in., and $Y = 0\cdot2\,\%$ compressive proof stress in tons/sq.in. The line was obtained by least squares, and the correlation coefficient for the thirty-seven results involved was 0·94, implying a significant relationship. The relation given was the best fit to the experimental data, but should not be extrapolated beyond the maximum static strength employed, because it is unlikely that the dynamic strength is ever less than the static strength, and, more probably, the ratio approaches asymptotically to unity at very high strength values.

COMPARISON OF DYNAMIC RESULTS WITH THOSE OBTAINED IN OTHER DYNAMIC EXPERIMENTS

Brown & Vincent (1941) of the National Physical Laboratory made high-speed tensile impact tests on a number of ferrous metals at rates of strain up to approximately 900 in./in./sec. Stress-strain measurements were made and showed that, in the case of a medium carbon steel and Yorkshire iron, a rise of yield stress accom-

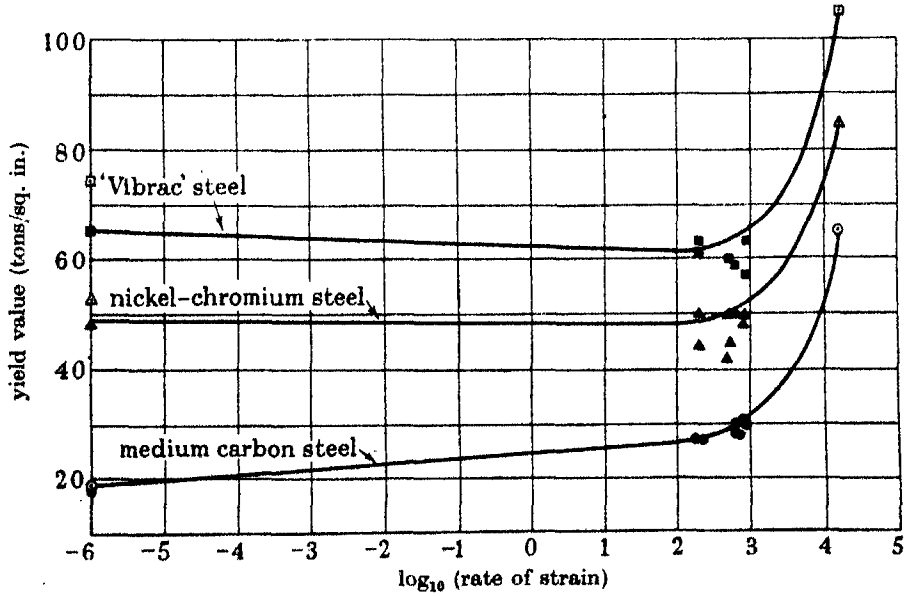


FIGURE 9. Effect of rate of strain on yield values of various steels.

material	tensile results	compression results
medium carbon steel	●	⊙
Ni–Cr steel	▲	△
'Vibrac' steel	■	⊡

Table 7. Results of tests on medium carbon, nickel chromium and 'Vibrac' steels

Details of materials taken from table 1, N.P.L. tensile results from table 2, and N.P.L. dynamic results from figures 13 and 14 of paper by Brown & Vincent, *Proc. Instn Mech. Engrs, Lond.*, (1941), **145** (3), 126–34

chemical analysis (%)								N.P.L. tensile results						results of dynamic tests			
															N.P.L. tensile tests		
C	Si	Mn	Ni	Cr	Mo	S	P	yield stress (tons/sq.in.)	ultimate stress (tons/sq.in.)	elongation on 4 √area (%)	reduction of area (%)	Izod test on 10 mm. × 10 mm. specimens (ft.lb.)	0·2 % proof stress (compression) (tons/sq.in.)	rate of strain (in./in./sec.)	yield stress (tons/sq.in.)	ultimate tensile stress (tons/sq.in.)	R.R.L. cylinder test dynamic comp. yield strength (tons/sq.in.)
medium carbon steel (normalized at 875° C)																	
0·30	0·20	0·67	0·10	0·09	—	0·570	0·046	18	35·0	39·5	60	40	19	180	27	35	—
														215	27	34	—
														590	30	35	—
														610	28	35	—
														620	28	34	—
														840	31	35	—
														890	30	37	—
														13,700	—	—	65
nickel-chromium steel (oil-hardened from 830° C, tempered at 600° C and quenched in water)																	
0·16	0·27	0·39	3·55	0·47	—	0·037	0·037	48	55·9	22	63	73	53	190	50	56	—
														190	44	53	—
														475	42	54	—
														500	50	59	—
														530	45	56	—
														680	50	59	—
														800	50	53	—
														840	49	55	—
														15,300	—	—	85
'Vibrac' steel (oil-hardened from 830° C, tempered at 600° C and cooled in air)																	
0·33	0·22	0·73	2·64	0·56	0·64	0·037	0·035	65·5	72·6	19	69	45	74	185	63	74	—
														190	61	71	—
														530	60	74	—
														580	59	70	—
														830	58	74	—
														880	63	73	—
														16,500	—	—	105

panied increased rate of strain. In the case of a low-carbon manganese steel, a nickel-chromium steel, and 'Vibrac' steel, there was a tendency for the yield stress to fall at first, but to increase to approximately its initial value at the higher rates of strain. The ultimate tensile stresses were not affected by the rate of strain.

Mr A. F. C. Brown has supplied samples of the medium carbon, nickel-chromium, and 'Vibrac' steels which he used in his tests, and their dynamic compressive yield strengths have been determined. Table 7 gives details of the chemical analyses of the materials investigated, taken from Brown & Vincent's paper, together with their mechanical properties, and the results of the present series of experiments. For all three materials, the dynamic compressive strength exceeded the 0·2 % static compressive proof stress, and also exceeded the ultimate tensile strength, whether measured statically or dynamically. Figure 9 shows the yield values obtained in both investigations plotted against the logarithm of the rate of strain; in every case a sharp rise of yield strength occurred at the high rates of strain employed in the method described in this paper. A part of this rise may perhaps be attributed to the fact that, in the present tests, the dynamic strength was determined in compression, whereas Brown & Vincent made tensile tests, and, in fact, in the static tests the 0·2 % proof stress, measured in compression, was higher than the tensile yield value in every case.

EFFECT OF PRE-STRAINING STEEL

Dr S. L. Smith, who, at the time the tests were made, was Superintendent of the Engineering Division of the National Physical Laboratory, suggested that tests be made to determine the dynamic compressive yield strength of strain-hardened mild steel. The material selected was a mild steel which, in the annealed condition, had a static tensile yield stress of 17 tons/sq.in., and a nominal ultimate tensile stress of 27·7 tons/sq.in. One set of specimens was made from bar in the annealed condition, a second set was made from bar which had been strained in tension to a true stress of 33·6 tons/sq.in. (nominal stress 26·2 tons/sq.in.), and a third set of specimens was made from bar strained in compression to an equal true stress. The results of the dynamic tests are given in table 8, the values of dynamic strength corresponding with the exact analysis having been obtained by correcting the values computed by the approximate formula.

The results given in table 8 indicate little difference between the dynamic strengths of specimens initially strained in tension and in compression, but both strengths were approximately 10 tons/sq.in. above that of the material in the annealed condition.

It is well known that the yield point of mild steel, tested under static conditions. may be raised by pre-straining of the type described. It was shown in figure 8 that the relationship between the dynamic and static strengths of all the steels tested appeared to depend solely on the static strength, and to be independent of the method by which the static strength was obtained. Consequently, in the present case, since pre-straining is likely to have produced an increase in the static yield point, the dynamic strength also would be expected to have increased. Increase of the 0·2 % proof stress from 16 to 20 tons/sq.in. would raise the dynamic compressive yield strength from 49 to 56½ tons/sq.in., so that the results given in table 8 appear to conform to the general trend of the results obtained with all types of steel.

TABLE 8. RESULTS OF DYNAMIC TESTS ON STRAIN-HARDENED MILD STEEL

			dynamic compressive yield strength	
condition of material	striking velocity (ft./sec.)	estimated rate of strain (in./in./sec.)	by approx. formula (tons/sq.in.)	by exact analysis (tons/sq.in.)
as annealed	1505	11,000	43	46
	1525	11,100	44	48
	1560	11,200	44	47
	1875	12,000	38	41
			mean =	45½
annealed and strained in tension	1190	11,000	58	62
	1450	11,800	51	55
	1565	11,700	47	51
	1870	12,900	49	52
	2210	14,000	44	47
			mean =	53½
annealed and strained in compression	925	8,900	52	58
	1395	12,000	53	57
	1395	12,100	56	60
	1935	13,800	56	59
			mean =	58½

TESTS ON DURALUMIN

Table 9 summarizes the results of tests on duralumin of various types, while the results are shown graphically in figure 8. In all cases the dynamic compressive yield strength was higher than the static proof stress, and increase of the static strength produced a decrease of the ratio of dynamic to static strengths. As in the tests on steel, there appeared to be a linear relationship between the ratio of dynamic to

TABLE 9. RESULTS OF TESTS ON DURALUMIN

			average values of			
				dynamic compressive yield strength		
Brinell hardness	0·2 % static proof stress Y (tons/sq.in.)	velocity range (ft./sec.)	rate of strain (in./in./sec.)	by approx. formula (tons/sq.in.)	by exact analysis S (tons/sq.in.)	ratio S/Y
57	7·0	1230–2270	11,800	16·5	17·8	2·54
66	8·15	445–1455	14,900	16·8	18·8	2·31
—	19·0	540–3120	14,300	29·0	31·0	1·63
118	24·3	1090–2100	15,600	32·0	33·9	1·40
133	25·85	495–1220	18,900	34·85	38·7	1·50

static strength, and the logarithm of the static strength (see figure 8 (*a*)). The best line through the experimental points was given by

$$S/Y = 4{\cdot}09 - 1{\cdot}89 \log_{10} Y, \tag{9}$$

where the symbols S and Y have the meaning employed in equation (8). As in the case of steel, the lines shown in figure 8 should not be extrapolated outside the range of the experiments.

TESTS ON SILVER, COPPER, LEAD, AND ARMCO IRON

Dynamic tests were made on standard silver having a diamond pyramid hardness of 123 to 147, but no information was available concerning the static values. Tests were also made on pure lead (99·97 %), electrolytic copper, and Armco iron, all these materials being normally very soft, and giving no definite indication of yield points in static compression tests. The results of the dynamic tests are summarized in table 10.

TABLE 10. RESULTS OF DYNAMIC TESTS ON SILVER, COPPER, LEAD, AND IRON

		average values of			
			dynamic compressive yield strength		
material	velocity range (ft./sec.)	rate of strain (in./in./sec.)	by approx. formula (tons/sq.in.)	by exact analysis (tons/sq.in.)	notes
standard silver	580– 775	10,800	32·7	37·6	V.D.H. = 123 } 7·5% Cu;
	300– 845	9,900	30·5	35·1	V.D.H. = 147 } 92·5% Ag
copper	330– 990	8,600	12·4	15·5	electrolytic
lead	210– 290	3,500	2·35	2·8	99·97 % pure
Armco iron	690–1000	14,700	43·5	49·6	composition: 0·016 % C; 0·006 % Si; 0·017 % S; 0·003 % P; 0·030 % Mn

Note. The silver, lead, and iron specimens were supplied by the Metallurgy Division of the National Physical Laboratory, together with the chemical analyses of the lead and iron.

In table 10 the results given for copper, lead, and iron are each the average of at least ten values which were in good agreement, and, as usual, they showed no variation with striking velocity. The value given by the pure iron is particularly interesting, because it is equal to that given by mild steel having a static strength of 18 tons/sq.in. (see table 5), whereas the static strength of the iron was not obtainable because no definite yield value could be found in compression tests.

COMPARISON OF THE VALUES OF X/L AND L_1/L GIVEN BY VARIOUS MATERIALS

Figure 10 shows X/L plotted against L_1/L for various materials, the lines given representing the best lines through the experimental points. The small numbers given against these lines are the striking velocities at which the values of X/L and

L_1/L were obtained. Superposed on this figure is the correction contour diagram given in figure 4 of part I, that is, the correction to be applied to values computed by the approximate formula to bring them into line with the exact analysis. The correction is small for hard materials such as armour steel, but, in general, increases when softer materials such as copper and lead are employed. Curve number 2 of figure 10 is interesting because it shows that 'normal' duralumin lies between hard

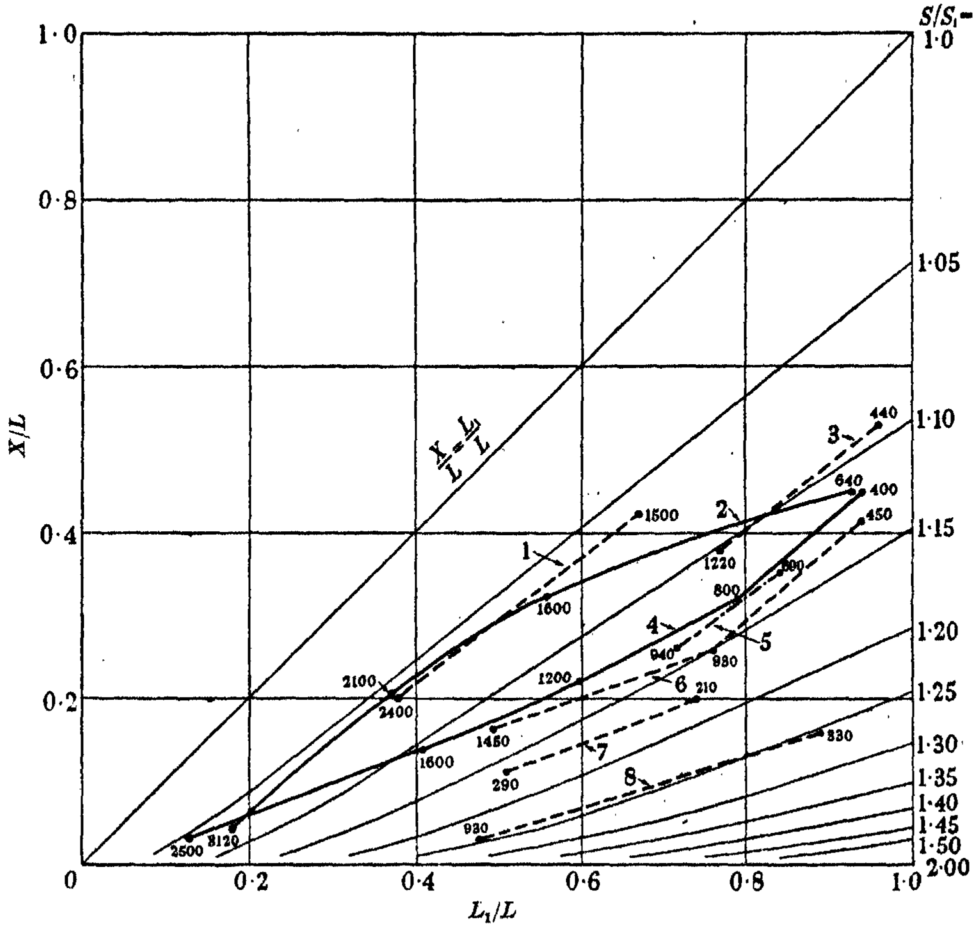


FIGURE 10. Variation of X/L with L_1/L for various materials.

ref. no.	material	static strength (tons/sq. in.)	average dynamic strength (tons/sq. in.)
1	armour plate	73	115
2	'normal' dural	19	31
3	hard dural	25·85	39
4	mild steel	18	49
5	Armco iron	—	50
6	soft dural	8·15	19
7	lead	—	2·8
8	copper	—	15·5

N.B. the small figures in the diagram show the striking velocity employed.

and soft duralumin at low velocities, but is similar to armour steel at velocities of 1600 to 2000 ft./sec. It will be appreciated that, although the values of X/L and L_1/L may be in agreement in such cases, so that the percentage corrections are equal, the actual values of the dynamic strength depend upon the density of the material and the striking velocity, and may therefore be very different.

It will be noticed from figure 10 that a specific value of X/L is given by different values of L_1/L, depending upon the material tested. For example, $X/L = 0{\cdot}18$ will be given by $L_1/L = 0{\cdot}35$ for armour steel and 'normal' duralumin, 0·5 for mild steel, 0·6 for soft duralumin, 0·7 for lead and 0·95 for copper. It would appear that these values of L_1/L represent the relationship between the ductilities of the materials at high rates of strain.

CONCLUSIONS

The following conclusions may be drawn from the experiments described:

(1) Using the method suggested in part I, it is possible to determine the dynamic compressive yield strength of various metallic materials at rates of strain estimated to be of the order of 10^4 in./in./sec.

(2) The dynamic compressive yield strengths may be computed from the experimental data by first using the approximate formula (22) of part I, and then adjusting the result to correspond with the exact equations (27) to (30) by applying a correction based on the dimensions X/L and L_1/L. This correction is of the order of 6 % for a hard material such as armour plate, and increases with decrease of the quality of the material, being of the order of 18 to 25 % for soft materials such as lead and copper.

(3) With the materials tested there was little change in the dynamic strength with variation of the striking velocity employed, probably because the rate of strain did not vary to any great extent with the striking velocity.

(4) The method was found to be inaccurate at high and low velocities; for instance, with mild steel, satisfactory results were obtained only within the range 400 to 2500 ft./sec. The range of velocities within which satisfactory results could be obtained varied with the quality of the material tested, soft metals giving results within a much lower range than that necessary for harder materials. Because of its failure at low velocities, the method cannot be used to bridge the gap between static and dynamic tests.

(5) Within the range of weapons available at the laboratory there was little change in dynamic strength with alteration of the specimen dimensions, probably because the corresponding change in rate of strain was not large.

(6) In general, the dynamic compressive yield strength S was greater than the static strength Y represented by the compressive stress giving 0·2 % permanent strain.

(7) For steels of various types, regardless of chemical composition and heat treatment, there was a relation between the ratio S/Y and the static strength Y, the ratio varying from approximately 3 when Y was 20 tons/sq.in. to 1 when Y was 120 tons/sq.in. The relation was given by the equation

$$S/Y = 5{\cdot}98 - 2{\cdot}42 \log_{10} Y.$$

(8) A similar relation was found for duralumin, for which

$$S/Y = 4{\cdot}09 - 1{\cdot}89 \log_{10} Y.$$

(9) An increase of dynamic strength was obtained by pre-stressing mild steel either in tension or in compression.

(10) Dynamic tests were made on the medium carbon steel, nickel-chromium steel, and 'Vibrac' steel which had been tested under tensile impact conditions by Brown & Vincent (1941), who found that, at strain rates up to 900 in./in./sec., there was an increase in the yield value of medium carbon steel, but little change in the case of the other two materials, while the ultimate tensile values were almost unaffected. The present dynamic tests gave increases of yield strength in all three cases.

(11) Dynamic compressive yield values were obtained for very soft materials such as pure lead, copper, and Armco iron, which, under static conditions, gave no definite yield values.

(12) A plot of the unstrained length X, expressed as X/L, versus the final overall length L_1, expressed as L_1/L, for the various materials showed that any specified value of X/L was associated with greater values of L_1/L for the more ductile materials, such as copper and lead, than for the more brittle materials, such as armour plate and duralumin. It appears, therefore, that the ductilities of materials under dynamic conditions may be compared by studying the ratio L_1/X for a specified value of X.

The author's thanks are due to the Director of the Road Research Laboratory and to the Chief Scientist, Ministry of Supply, for permission to use the data contained in this paper.

REFERENCE

Brown, A. F. C. & Vincent, N. D. G. 1941 *Proc. Instn Mech. Engrs, Lond.*, **145** (3), 126.

The use of flat-ended projectiles for determining dynamic yield stress

III. Changes in microstructure caused by deformation under impact at high-striking velocities

By W. E. Carrington and Marie L. V. Gayler, D.Sc.

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 24 *July* 1947)

[Plates 1 to 8]

The mechanism of deformation of metals at high velocities has been studied by examining the microstructure of cylinders of mild steel, duralumin and standard silver (of composition Ag 92·5 %, Cu 7·5 %) of similar hardness, and of steel balls after impact on steel plates at velocities from 300 to 3830 ft./sec.

The means by which the stress on impact was relieved depended on the material, and was first the formation of twins or 'compression bands', i.e. by block movement of wedges of material within individual grains, or by cracking. When the applied stress could no longer be relieved in this way, plastic deformation occurred.

The amount of residual strain in deformed mild steel and duralumin has been examined by observing changes of microstructure after annealing.

Hardness surveys were made on longitudinal sections of the cylinders and the results correlated with the microstructures and with observations of the limits of strain due to impact.

Introduction

A number of investigations on the deformation of metals at high speeds was carried out during the war on behalf of the Armour Piercing Projectiles Sub-Committee of the Ordnance Board under the Chairmanship of Dr C. Sykes, F.R.S., and in some cases metallographic examination of the specimens gave interesting information on the nature of deformation at high speeds. This paper describes the results of the examination of some small cylinders of mild steel, duralumin and standard silver (composition 92·5 % silver, 7·5 % copper) and of some hardened steel balls, which had been fired with various striking velocities against steel plates. The duralumin and standard silver cylinders were aged to the approximate hardness of the mild steel cylinders.

Metallographical examination showed that the response to impact occurred in different ways, depending on the nature of the material from which the cylinders had been made. The results obtained were correlated with hardness measurements and threw considerable light on the mechanism by which deformation under impact at high velocities can occur.

Experimental

The relevant data regarding the specimens examined are given in table 1. The mild steel cylinders were cut from rolled plate and the duralumin and standard silver ones from rolled bar. Both sets of cylinders were cut out of the rolled material so that the axis of the cylinder was parallel to the direction of rolling. The duralumin cylinders were subsequently solution heat-treated and aged at room temperature and the standard silver cylinders, also solution heat-treated, were aged to about the same hardness by ageing for 10 min. at 240° C.

TABLE 1

specimen		condition	diameter (in.)	length (in.)	V.D.P.H.* (HD/10)	striking velocity (ft./sec.)	unstrained length (in.) R.R.L.†	unstrained length (in.) N.P.L.‡
mild steel	no. 13	as rolled	0·350	0·500	130	510	0·23	0·18
mild steel	O_1	as rolled	0·500	1·005	130	1110	0·24	0·27
mild steel	O_2	as rolled	0·500	1·002	130	1250	0·18	0·21
mild steel	O_6	as rolled	0·500	0·999	130	2140	0·07	—
duralumin	D_3	aged room temperature	0·360	0·500	125	500	0·24	0·24
duralumin	D_4	aged room temperature	0·246	1·021	125	640	0·45	0·6
duralumin	D_{71}	aged room temperature	0·311	1·007	125	1145	0·46	0·45
duralumin	D_7	aged room temperature	0·311	1·000	125	2160	—	—
standard silver	B_2	aged 10 min. at 240° C	0·358	0·500	142	300	0·215	—
steel ball	no. 25	Hoffmann	0·125	—	—	1975	—	—
steel ball	no. 20	Hoffmann	0·125	—	—	3830	—	—

* V.D.P.H. = Vickers diamond pyramid hardness. The values in this column were those obtained from the Vickers pyramid hardness testing machine using a load of 10 kg. for 15 sec. on a diamond pyramid indenter of angle 136°.

† Measured by determining point at which diameter increased by 0·001 in.

‡ Measured by determining the point at which Neumann lamellae or compression bands first appeared.

The size of cylinders fired with striking velocities of (i) 600 ft./sec. or less, was 0·35 in. diameter, length 0·5 in., and (ii) above 600 ft./sec., 0·31 in. diameter, length 1·0 in.

Hoffmann steel ball-bearings, $\frac{1}{8}$ in. diameter, were also fired.

Longitudinal and medial sections of the cylinders were polished and suitably etched, after which macro- and micro-examinations were carried out. Vickers diamond hardness impressions were subsequently made over the specimens after micro-examination.

(1) *Mild steel cylinders*

Figures 1, 2 and 3, plate 1, show the macrostructure of mild steel cylinders fired with striking velocities of 1110, 1250 and 2140 ft./sec., under magnifications of 5 diameters. The structures were developed by Fry's reagent (Berglund 1931), i.e. a strongly acid solution of copper chloride. It is clear that plastic flow had taken place in the head of the cylinder, the degree depending on the striking velocity, and that tensional forces must have come into play along the flanges of the head (figure 3).

The microstructure developed by re-polishing and etching with 5 % picric acid and ½ % nitric acid in alcohol, differed from that of mild steel deformed at a more ordinary speed, principally in the large number of Neumann lamellae observed. They were formed in scattered grains in which no visible plastic deformation was detected, i.e. in the region near the limit of strain, as shown in figure 4, plate 2; and, as the deformed head of the cylinder was approached, the Neumann lamellae increased in number and became increasingly deformed. The deformation of the lamellae was similar to that shown in figure 29, plate 7.

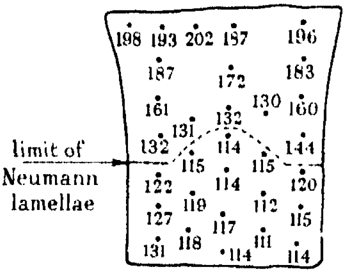


FIGURE 6. Mild steel cylinder. Striking velocity 510 ft./sec. Magn. × 3.

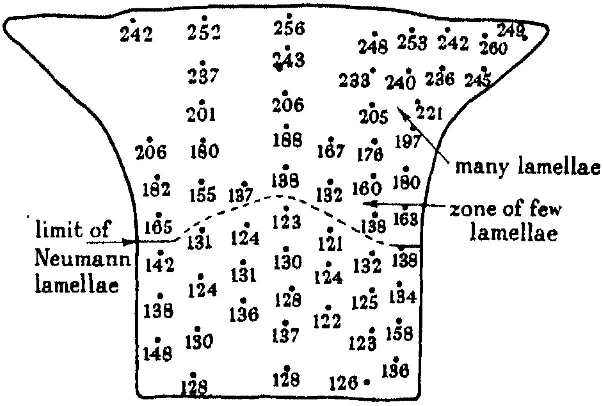


FIGURE 7. Mild steel cylinder. Striking velocity 1110 ft./sec. Magn. × 3.

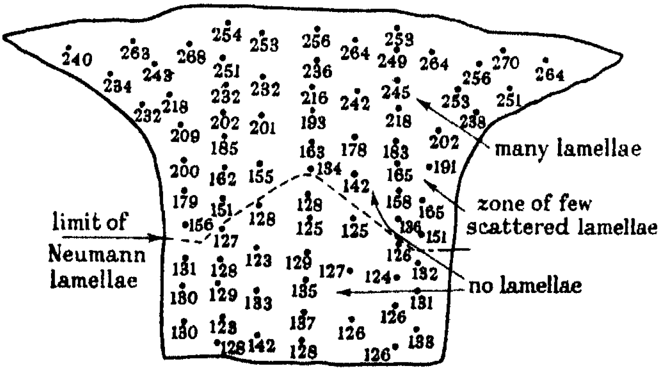


FIGURE 8. Mild steel cylinder. Striking velocity 1250 ft./sec. Magn. × 3. (N.B. Figures 6 to 8 are plans of V.D.P.H. impressions.)

These facts show that Neumann lamellae must have been formed on impact *before* plastic deformation occurred.

It is of interest to note that a narrow layer of material occurred immediately below the centre of the head of the cylinder, which was less heavily deformed than material at a greater depth (figure 5, plate 2).

As would be expected, maximum hardening occurred where deformation was greatest, as shown in figures 6, 7 and 8, which are diagrams of cylinders fired with striking velocities of 510, 1110 and 1250 ft./sec., respectively. The broken line in each diagram represents the limit of strain above which, and up to the head of the cylinder, Neumann lamellae appear, first in very small numbers, i.e. one or two in isolated crystals and then, in increasing numbers, in an increasing number of crystal grains as the head is reached. It is of interest to note the cone-shaped region of unstrained material which penetrates into the deformed portion of the cylinder, as shown in figures 6, 7 and 8. The extent of penetration appears to depend upon the striking velocity.

(2) *Steel balls*

Figures 9 and 10, plates 2 and 3, show the etched sections of two hardened steel balls (Hoffmann) fired with striking velocities of 1975 and 3830 ft./sec., respectively, against mild steel plate of Vickers diamond hardness value about 140. The former specimen shows the presence of cracks and no plastic deformation; the latter has been considerably plastically deformed, as well as cracked. It is significant that (i) cracks alone have been formed on impact in the ball fired with the lower striking velocity and (ii) the direction of the cracks in the second ball have been changed owing to subsequent plastic deformation. This behaviour indicates that the first result of impact at high velocity is the formation of cracks (cf. figures 9 and 10).

The white 'lines' shown in figures 10 and 11, plate 3, are probably due to regions of maximum shearing stress, where the heat generated locally has, most probably, raised the temperature of the material to the neighbourhood of the melting-point.

(3) *Duralumin cylinders*

A different type of deformation structure occurred in the duralumin cylinders fired at high striking velocities.

On etching with Keller's reagent (Keller 1935), what have been termed 'compression bands' were revealed. Figures 12 and 13, plate 3, show these bands in the heads of cylinders fired with striking velocities of 500 and 640 ft./sec., respectively. Some of the crystal grains in the matrix are free from such bands, as shown in figure 12 and figure 14, plates 3 and 4. Plastic deformation, to a greater or less degree depending on the striking velocity, also occurs in the heads of the cylinders, as shown in figure 15, plate 4, in the case of the cylinder fired with a striking velocity of 2160 ft./sec. In this particular specimen the heat generated on impact at the extreme head of the cylinder had been so great, that a local rise in temperature had occurred which resulted in a complete change in microstructure, as shown in figure 16, plate 4. This structure merges into that shown in figure 15, i.e. deformed 'compression' bands; the latter deformation decreasing as the unstrained region is approached. The heat generated at the extreme head of the cylinder has caused a change in microstructure similar to that associated with the accelerated ageing of duralumin due to cold-working and subsequent ageing at a high temperature. It is, therefore, deduced that the maximum temperature attained in the head of the cylinder must have been at least 200° C and very probably higher. Hardness measurements revealed that hardening had occurred to an extent comparable with the maximum it is possible to attain with cold-worked, heat-treated duralumin, i.e. 150 to 160 V.D.P.H.* (see figures 21 to 23).

---

* V.D.P.H. = Vickers diamond pyramid hardness; see table 1.

Reference must now be made to the nature of the 'compression bands'. These are not 'twin formations' since they do not differ in orientation from the surrounding matrix when examined under dark-ground illumination. They are also not 'slip bands' since with similar oblique illumination from opposite sides, both edges of the bands become illuminated, as shown in figures 18 and 19, plate 5. If these two photomicrographs are examined, it will be seen that the edges of bands (white) in one photomicrograph differ in position from those in the identical area in the other; a pair thus indicating the width of the 'compression band'. This fact also shows that 'compression bands' must have been selectively attacked by the etching reagent in comparison with the matrix, as shown diagrammatically in figure 20.

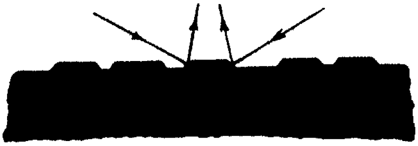


FIGURE 20. Diagram showing cross-section of compression bands, examined under oblique light.

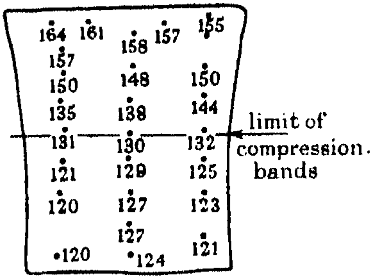


FIGURE 21. V.D.P. hardness survey of duralumin cylinder. Striking velocity 500 ft./sec. Magn. × 3.

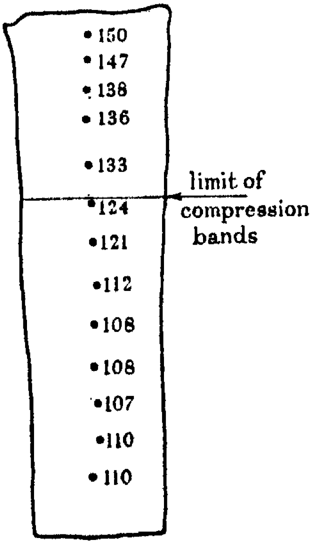


FIGURE 22. V.D.P. hardness survey of duralumin cylinder. Striking velocity 640 ft./sec. Magn. × 3.

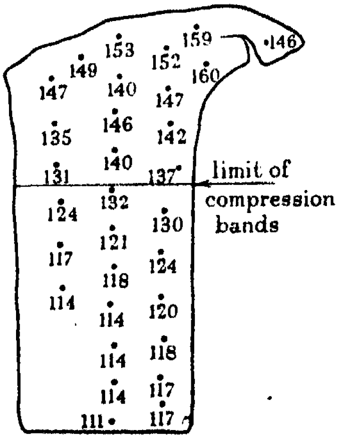


FIGURE 23. V.D.P. hardness survey of duralumin cylinder. Striking velocity 1145 ft./sec. Magn. × 3.

'Compression bands' are considered to be wedges of material which have moved in a direction relatively parallel to each other in individual grains (figures 12 to 14). The individual crystal has more or less crumpled up under impact, like a concertina (figures 12 to 14); the axis of each crystal remaining unchanged with respect to those of its neighbours. The non-appearance of 'compression bands' in certain crystals is probably due to the relative orientation of the grain with respect to the axis of compression. The plane of polish of these particular grains is probably parallel, or nearly so, to the direction of slip, i.e. (110) planes.

Barrett & Levenson (1940) observed what they also have termed 'compression bands' in aluminium under static compression. These, in contrast with the above, are on a macroscopic scale and of the same order of size as Lüder's lines. These authors, however, found by X-ray analysis that the bands were *not* twin formations.

In order to determine whether the formation of 'compression bands' was accompanied by any distortion between the area of contact and the surrounding material, the duralumin cylinder fired with a striking velocity of 640 ft./sec. (figure 22) was heat-treated for 1 hr. at 500° C. Recrystallization occurred throughout the strained area with absence of any exaggerated grain growth. The grains in the very heavily worked area at the extreme head of the cylinder were finer than those immediately below, which were generally uniform in size. Similarly, the duralumin cylinder fired with a striking velocity of 2160 ft./sec. (figures 15 and 16) was annealed at 500° C but for the longer period of 1 day. Figure 24, plate 5 shows the microstructure revealed by etching. The grain size was more or less uniform in the area in which recrystallization had occurred, i.e. in the strained area. Again, no region of exaggerated grain growth was found; a clear line of demarcation existing between the recrystallized grains and the original unstrained material (figure 24).

These results show that where no visible plastic deformation was observed, the residual strain caused by the formation of 'compression bands', between the latter and the surrounding material, must have been below the critical amount which would have given rise to excessive grain growth (Carpenter & Elam 1920). This residual strain may, however, be responsible for the differential etching effect between a 'compression band' and adjacent material (figure 20).

As in the case of the formation of Neumann lamellae, 'compression bands' appear to be formed *before* plastic deformation takes place. Figure 17 shows an area in the head of the cylinder fired with the striking velocity of 1145 ft./sec. The stepped edges of the bands and the lines across them, revealed by differential etching owing to the presence of areas of worked material, are evidence that 'slip' occurred *after* the 'compression bands' had been formed. Figure 15 shows general distortion of the 'compression bands' in the head of a cylinder fired with a striking velocity of 2160 ft./sec.

The results of hardness measurements made on cylinders fired with striking velocities of 500, 640 and 1145 ft./sec. are shown diagrammatically in figures 21 to 23. The limit of 'compression bands' is found to be approximately coincident with the limit of strain as determined by hardness measurements.

(4) *Cylinders of standard silver (composition 92·5 % silver, 7·5 % copper)*

These silver cylinders were aged to a slightly higher hardness value than that of mild steel. The polished sections were etched by the reagent used by Norbury (1928) and the presence of another type of deformation figure was observed. Cylinders fired with striking velocities of 300 to 800 ft./sec. and over, were characterized by the appearance of narrow 'bands', with pointed ends, of differing lengths and thickness in many of the crystals, see figure 25, plate 6. These were most numerous in grains in the extreme head of the cylinders but, even there, some grains were devoid of these 'bands' (probably for the same reason as mentioned in regard to absence of 'compression bands' in duralumin cylinders).

It is of interest to note the course of these 'bands' when a twin is present in the crystal grain. Figure 26, plate 6, shows such a grain (right-hand side); the 'bands' extend from the crystal boundary to one side of the twin and then re-appear on the opposite side of the twin. Examination of these 'bands' under oblique illumination leads to the conclusion that they are probably twinned regions and not 'compression bands'.

Slip, however, is superimposed on these 'bands' in certain regions, particularly in those in the extreme head of the deformed cylinder. Figures 27 and 28, plate 6, are photomicrographs of two such areas. Similarly, the curvature of the 'bands' in figure 25 may also be due to slight plastic deformation.

It was not possible at the time to check by X-ray analysis whether these 'bands' were twinned areas, even if it were possible on such fine-grained material. Apart from the evidence above, it is highly probably that twins have been formed, since standard silver twins readily. Further evidence was, however, obtained microscopically. Figure 28 shows a grain boundary (top) which is serrated in a manner associated with twinning and not that associated with slip. Hence, the parallel lines shown below the boundary (figure 28) are the edges of twins which are due to 'block-movement' of comparatively large portions of the grain.

(5) *Examination of 'block-movement' in mild steel cylinders*

In order to determine whether 'block-movement' in the mild steel cylinders takes place without leaving any significant residual distortion, the following experiments were carried out:

(i) A piece of mild steel was selected, in which a hole had been made by an explosive charge. In the neighbourhood of the hole, numerous deformed Neumann lamellae were present in plastically deformed grains, as shown in figures 29 and 30, plate 7, under magnifications of 200 and 500 diameters, respectively. A specimen cut from this area, i.e. figure 30, was heated for ¼ hr. at 650° C, and then lightly repolished and etched. As shown in figure 31, plate 7, which is the same area as figure 30, complete recrystallization has occurred and the Neumann lamellae have disappeared. This fact is due to recrystallization of the plastically deformed crystal grains and distorted Neumann lamellae, followed by grain growth.

(ii) Further away from the hole in the same piece of mild steel, a region existed in which no visible plastic deformation could be detected but only comparatively few Neumann lamellae, as shown in figure 32, plate 8. A specimen was cut from the portion and heat-treated for 5 min. at 750° C *in vacuo*. The microstructure of a marked place before and after heat-treatment is shown in figures 33 and 34, plate 8, respectively, under magnifications of 500 diameters. Apart from the disappearance of the pearlite and diminution in size of the Neumann lamellae, due, probably, both to the repolishing process and to absorption as a result of heat-treatment, no change in microstructure was observed, i.e. no recrystallization. Hence, no significant residual distortion can have existed between the Neumann lamellae and the adjacent material else recrystallization would have occurred.

Hence, it is deduced that in areas where no visible plastic deformation of Neumann lamellae is observed, the latter have been formed with no significant residual distortion.

DISCUSSION

The facts described above show that, under impact at high velocity, relief of stress takes place by two processes, the first followed by the second; i.e. by the formation of either twins, compression bands or cracks, according to the nature of the material in question; and (ii) by subsequent plastic deformation. In brief, when the grain can no longer adapt itself to relieve the applied stress by one of the above processes, i.e. by the formation of twins or compression bands, resistance to slip breaks down and plastic deformation sets in. This initial resistance to slip takes place in individual grains by a 'block' movement of material within the grains, i.e. a wedge (or wedges) of material moves relatively to the rest of the grain, thereby relieving the stress.

When twinning results from this 'block' movement no residual distortion occurs; but when 'compression bands' are formed, it is probable that the shifting of wedges of material may cause slight residual distortion. This 'block' movement is, therefore, a *local* movement in *individual* grains, which occurs initially in relief of stress applied with high velocities. In contrast, plastic deformation, or slip, is the general relief of stress by all the grains concerned. This takes place *after* the 'block' movement of individual grains has occurred.

The occurrence of 'block movement' prior to the general plastic deformation may be associated with the higher yield point observed when metals are stressed under dynamic conditions (Brown & Vincent 1941).

The work described above was carried out in the Metallurgy Division of the National Physical Laboratory on behalf of the Director General of Scientific Research (Defence), by whose permission this paper is published.

REFERENCES

Barrett, C. S. & Levenson, L. H. 1940 *Proc. Instn Mech. Engrs, Lond.*, **137**, 112.

Berglund 1931 *Metallographer's Handbook of Etching*, p. 59. London: Sir Isaac Pitman & Sons.

Brown, A. F. C. & Vincent, N. D. G. 1941 *Proc. Instn Mech. Engrs*, **145** (3), 126.

Carpenter, H. C. H. & Elam, C. F. 1920 *J. Inst. Metall.* **24**, 83.

Keller, F. 1935 *Min. & Metall.* **16**, 513.

Norbury, A. L. 1928 *J. Inst. Metall.* **39**, 145.

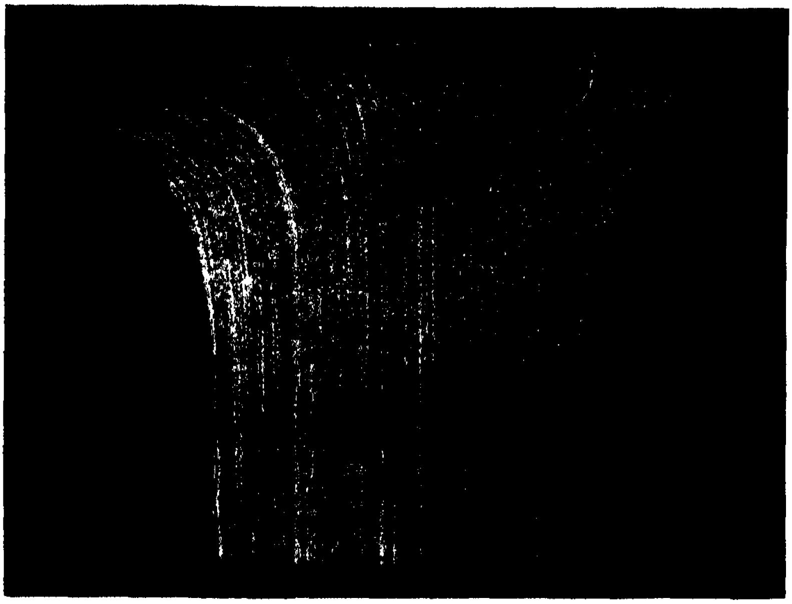


FIGURE 1

FIGURE 2

FIGURE 3

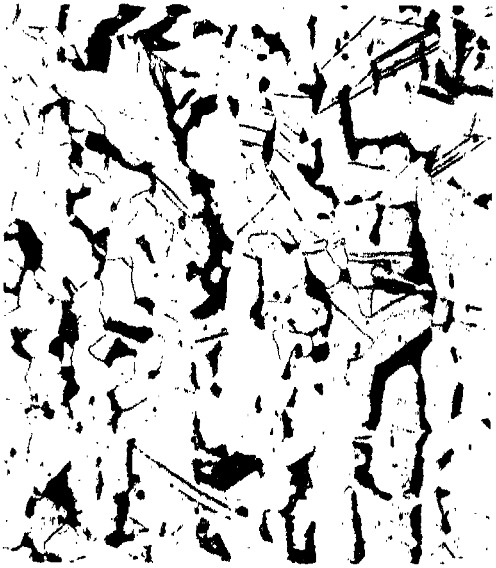


FIGURE 4

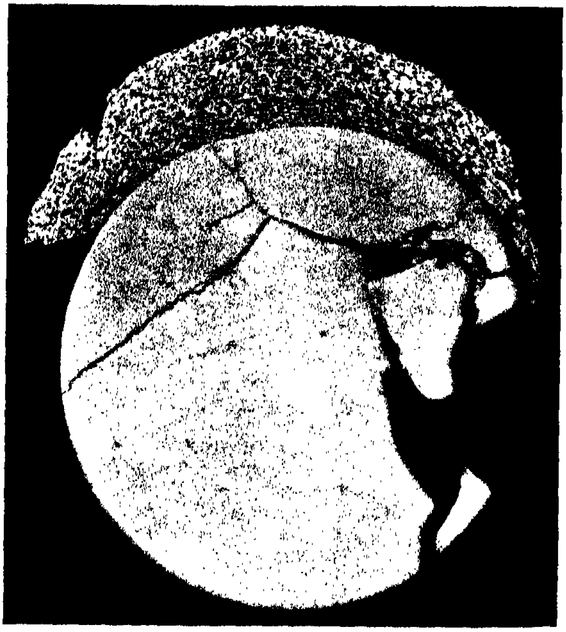


FIGURE 9

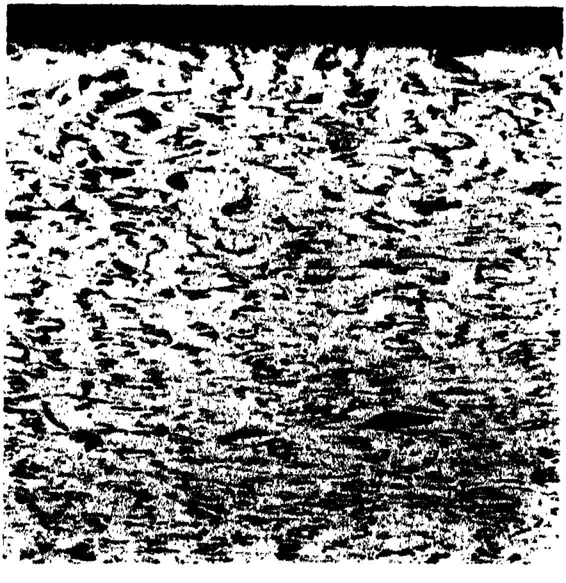


FIGURE 5

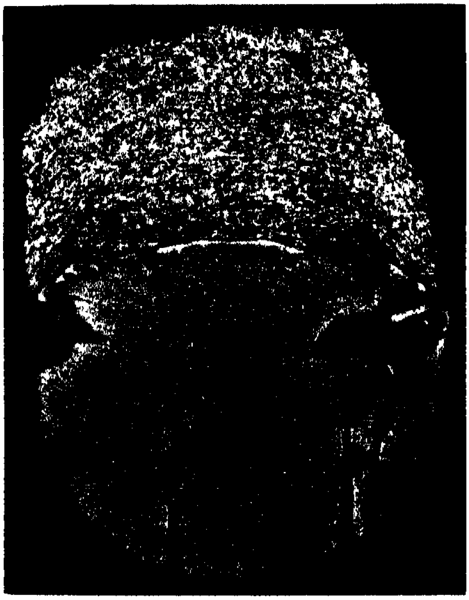


FIGURE 10

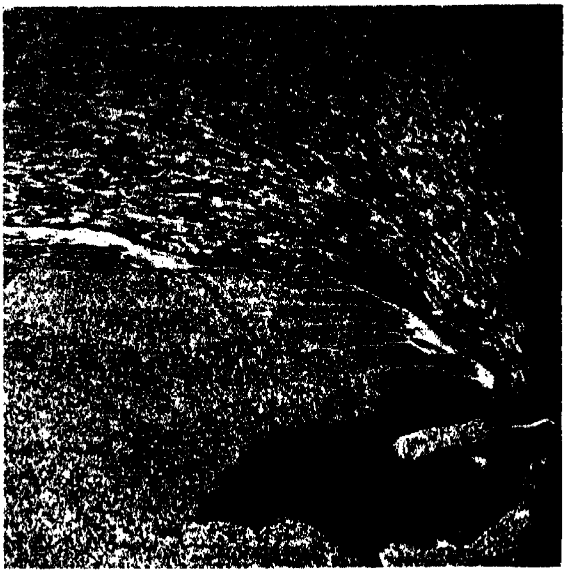


FIGURE 11

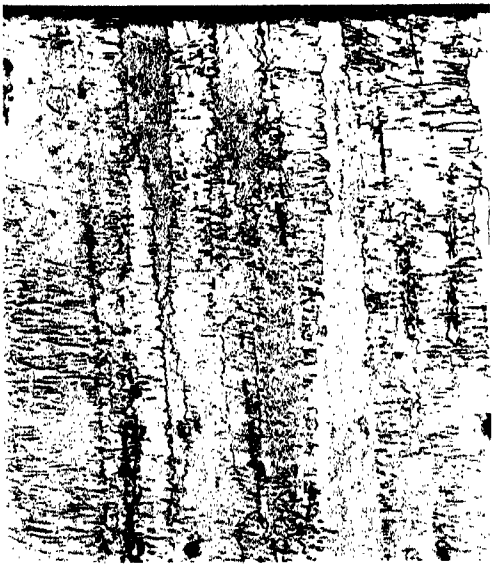


FIGURE 12

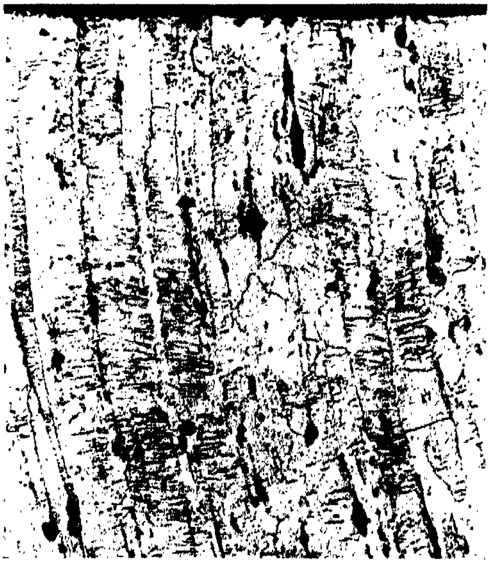


FIGURE 13

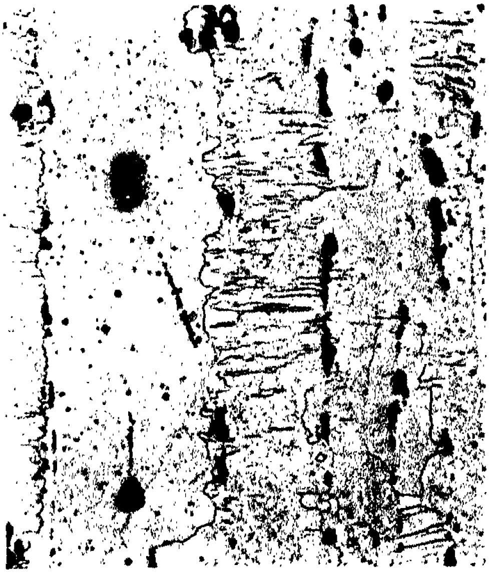


FIGURE 14

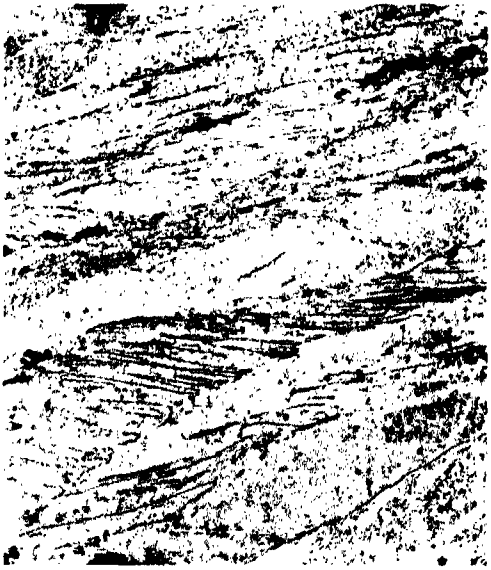


FIGURE 15

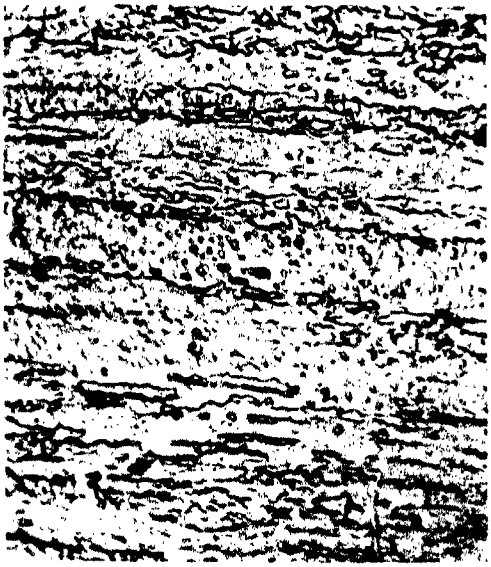


FIGURE 16

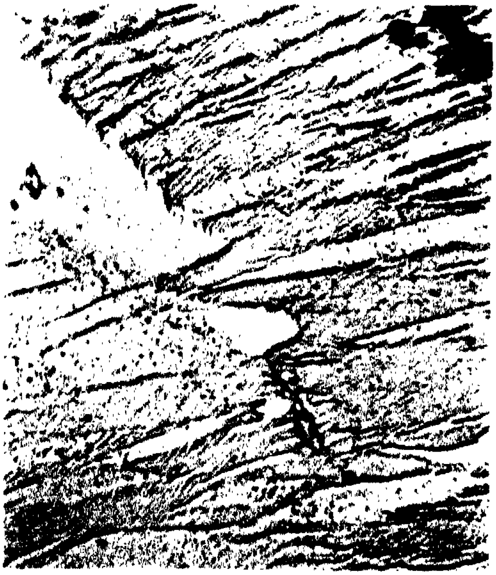


FIGURE 17

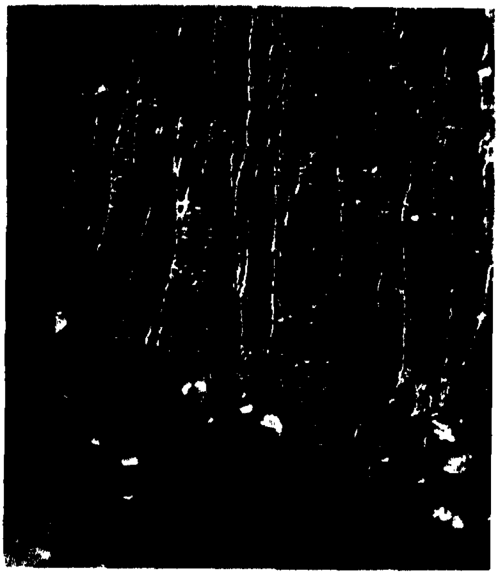


Figure 18

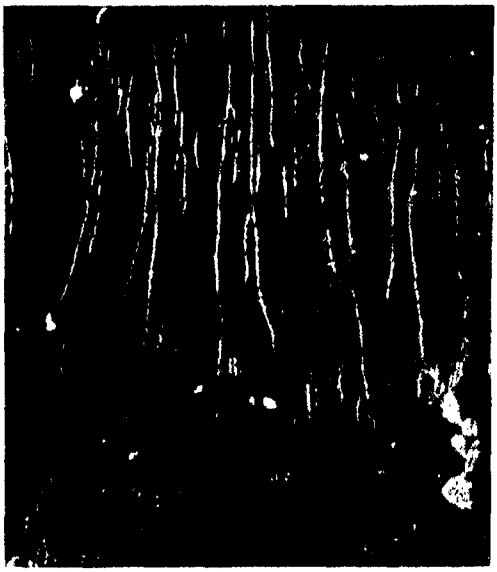


Figure 19

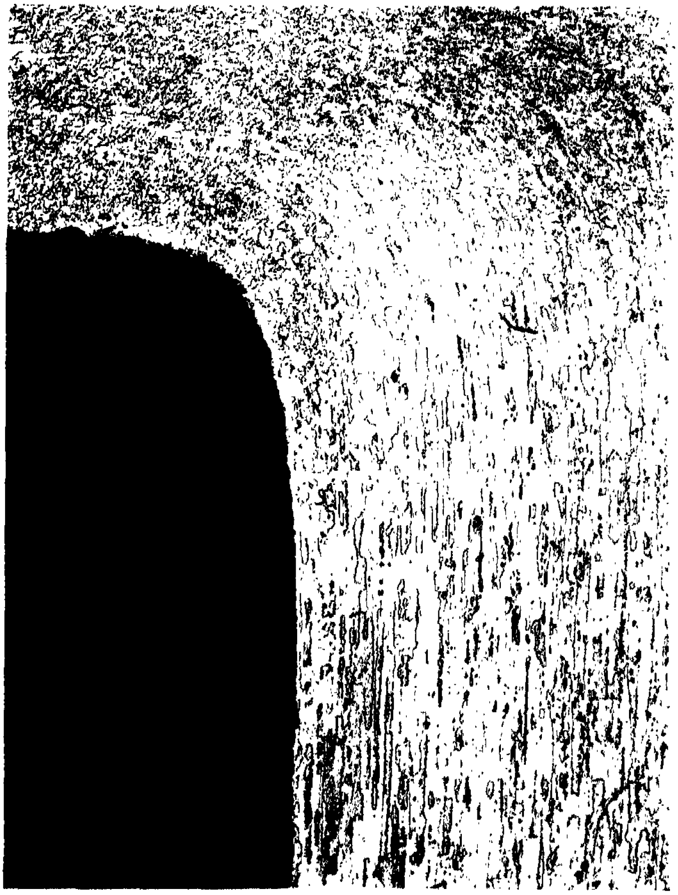


Figure 24

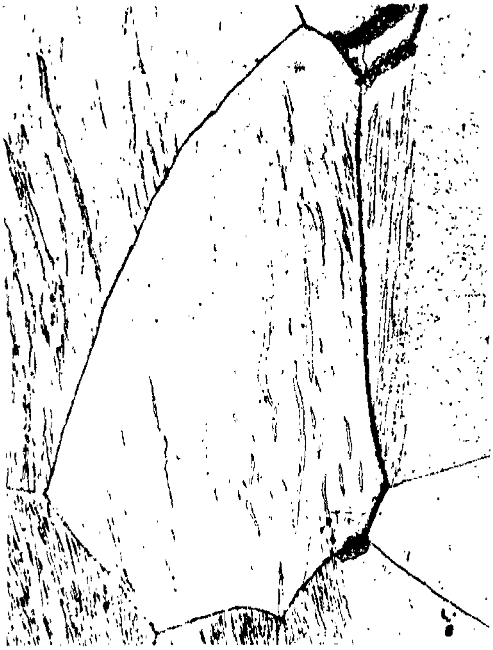


FIGURE 25

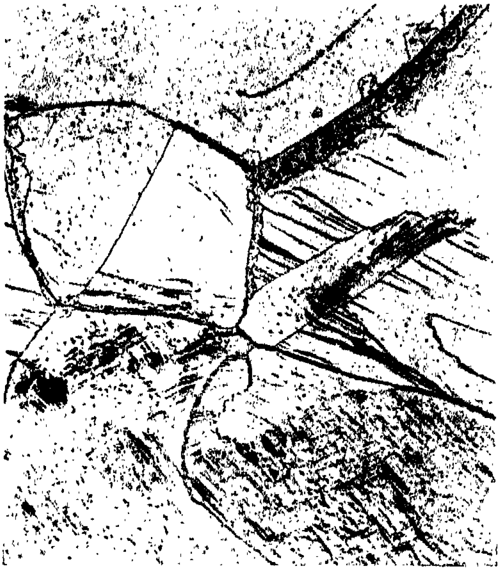


FIGURE 26

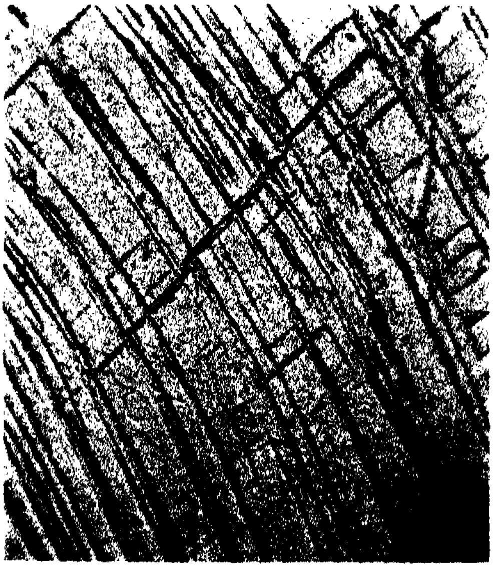


FIGURE 27

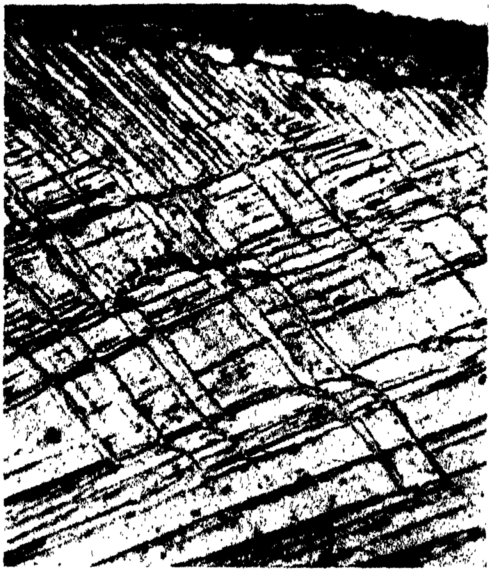


FIGURE 28

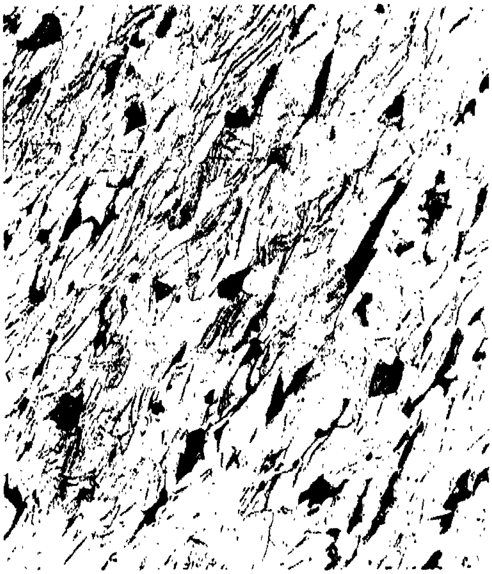


FIGURE 29

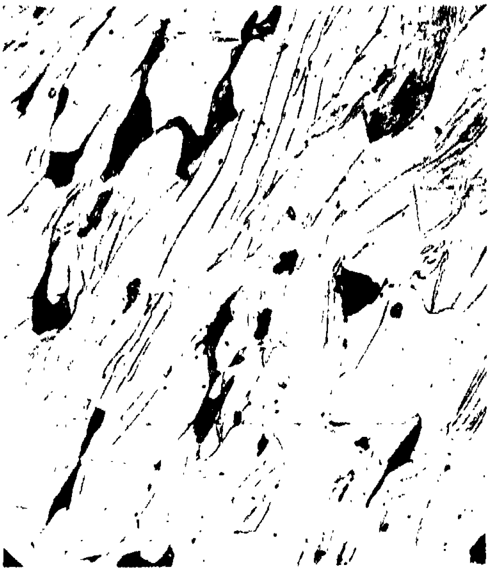


FIGURE 30

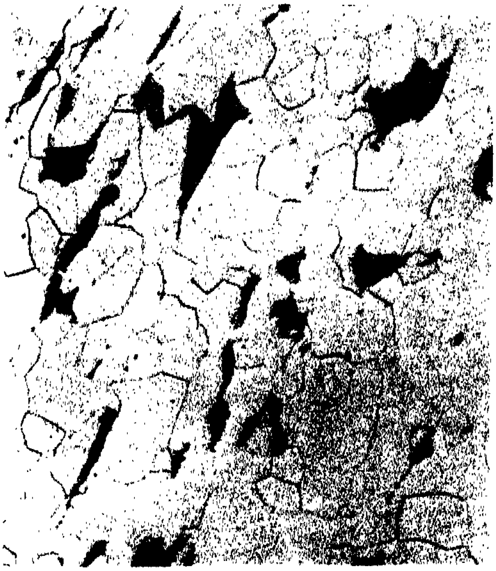


FIGURE 31

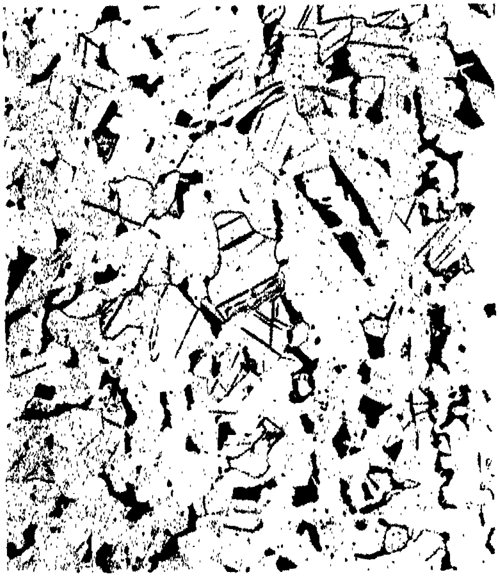


FIGURE 32

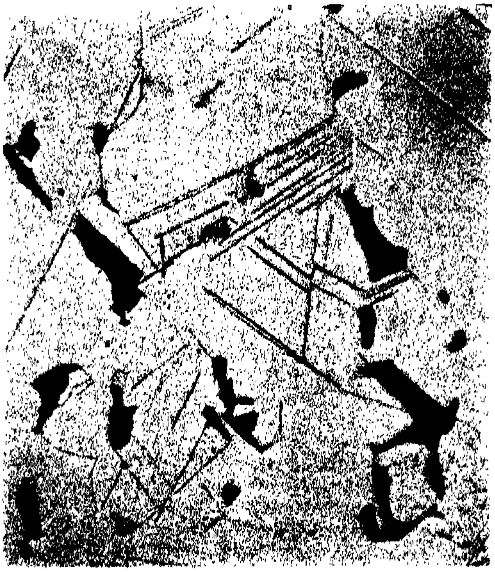


FIGURE 33

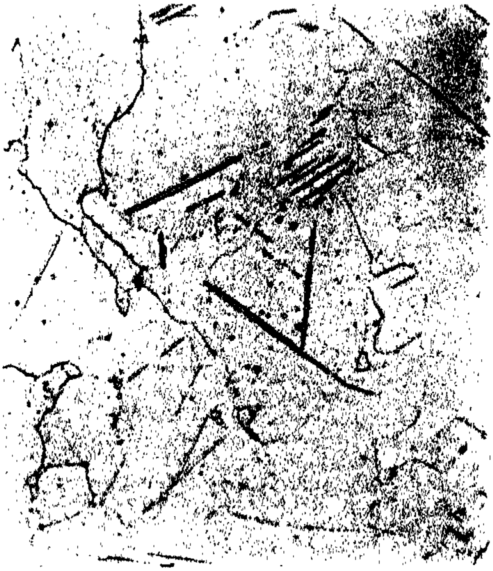


FIGURE 34

DESCRIPTION OF PLATES 1 TO 8

PLATE 1

FIGURE 1. Mild steel cylinder. Striking velocity 1110 ft./sec. Fry's micro etch. Magn. × 4.

FIGURE 2. Mild steel cylinder. Striking velocity 1250 ft./sec. Fry's micro etch. Magn. × 4.

FIGURE 3. Mild steel cylinder. Striking velocity 2140 ft./sec. Fry's micro etch. Magn. × 4.

PLATE 2

FIGURE 4. Mild steel cylinder. Striking velocity 510 ft./sec. Etched. Magn. × 200.

FIGURE 5. Head of mild steel cylinder. Striking velocity 1250 ft./sec. Etched. Magn. × 120.

FIGURE 9. Steel ball no. 25. Striking velocity 1975 ft./sec. Magn. × 24.

PLATE 3

FIGURES 10 and 11. Steel ball no. 20. Striking velocity 3830 ft./sec. Etch – 4 % nitric acid in alcohol. Magn. (10) × 24, (11) × 80.

FIGURE 12. Head of duralumin cylinder. Striking velocity 500 ft./sec. Keller's etch. Magn. × 120.

FIGURE 13. Head of duralumin cylinder. Striking velocity 640 ft./sec. Keller's etch. Magn. × 80.

PLATE 4

FIGURE 14. Head of duralumin cylinder. Striking velocity 640 ft./sec. Magn. × 240.

FIGURE 15. Head of duralumin cylinder. Striking velocity 2160 ft./sec. Magn. × 400.

FIGURE 16. Another area in head of duralumin cylinder shown in figure 15. Magn. × 1600.

FIGURE 17. Head of duralumin cylinder. Striking velocity 1145 ft./sec. Magn. × 800. (Figures 14 to 17 all specimens etched in Keller's reagent.)

PLATE 5

FIGURES 18 and 19. Duralumin cylinder. Striking velocity 2160 ft./sec. Appearance of compression bands in head under oblique lighting from opposite directions. Magn. × 720.

FIGURE 24. Duralumin cylinder. Striking velocity 2160 ft./sec. Annealed 1 day at 500° C. Etched in Keller's reagent. Magn. × 20.

PLATE 6

FIGURES 25 to 28. Standard silver cylinder. Striking velocity 300 ft./sec. Norbury's etching reagent. Magn. (25 and 26) × 120, (27 and 28) × 1600.

PLATE 7

FIGURE 29. Mild steel plate. $C = 0{\cdot}15\,\%$. Magn. × 160.

FIGURE 30. Area in figure 29. Magn. × 400.

FIGURE 31. Area in figure 30 after annealing for 15 min. at 650° C. Magn. × 400.

PLATE 8

FIGURE 32. Mild steel plate, T_3B, as received. Magn. × 120.

FIGURE 33. Area in figure 32. Magn. × 400.

FIGURE 34. Area in figure 33 after annealing for 5 min. at 750° C. Magn. × 400. (Figures 29 to 34 etched in equal parts of $\frac{1}{2}\,\%$ nitric acid and 5 % picric acid in alcohol.)

The determination of the masses necessary to produce a given gravitational field

By E. C. Bullard, F.R.S. and R. I. B. Cooper

Department of Geodesy and Geophysics, University of Cambridge

(*Received* 6 *January* 1948)

Two methods are described for calculating the distribution of mass at a given depth that will produce a specified gravitational field. In one the observed field is extrapolated downwards by a finite difference procedure; in the other a solution is obtained in the form of an integral which can be evaluated numerically. Tables of the quantities necessary for the practical use of the method are provided, and real and artificial examples are given. The limit to the possible depth of the masses producing an anomaly is also discussed.

1. STATEMENT OF THE PROBLEM

A gravity survey yields values of the gravitational field over the surface of the earth. By the Bouguer reduction allowance may be made for the varying heights of the points of observation, for the attraction of the ellipsoid to which the earth approximates and for the attraction of topographic irregularities. The residue, the Bouguer anomaly, represents the attraction of excesses and defects of density within the earth. The 'geological interpretation of gravity anomalies' consists largely of estimating the magnitudes and positions of these masses, and determining their relation to the structure of the area.

The usual procedure is to guess a suitable mass distribution and then to calculate the gravitational field it produces and to compare this with the observations. It is then possible to adjust the masses by a process of trial and error until a good fit is obtained. Much labour and ingenuity has been expended on the construction of tables and diagrams to facilitate such computations (Cassinis, Dore & Ballarin 1937; Meinesz 1934; Numerov 1927; Jung 1927). In comparison, curiously little attention has been given to the direct calculation of the masses from the anomalies. In this paper we develop methods for this purpose and apply them to certain examples. We deal throughout with the gravitational problem, but closely analogous methods could be used in other branches of potential theory.

2. INDETERMINACY OF THE PROBLEM

The problem is: given, over a plane, the normal component g of a field derivable from a potential, to find the mass distribution producing the field. In practice the field g is usually the Bouguer anomaly and the plane over which it is known is horizontal; for verbal convenience we assume this to be so throughout.

The problem is clearly indeterminate, for, by a theorem due to Green, any mass can be replaced by a suitable distribution at any shallower depth without affecting the field at the surface. The problem may be rendered more nearly determinate by seeking a distribution at a specified depth. A family of such solutions for different depths gives a good view of the whole set of possible solutions. By forming linear combinations of its members, further distributions may be obtained though these do not include all those possible.

The problem is still not completely determinate, for it is always possible to add to any solution a simple harmonic mass distribution of arbitrarily great amplitude, provided that the wave-length of the added distribution is sufficiently short. If $A \sin px$ is the added mass distribution at a depth z, where x is a co-ordinate measured horizontally, then the field it produces at the surface is

$$2\pi kA\, e^{-pz} \sin px,$$

where k is the constant of gravitation. This can be kept less than any quantity, however small, if p is made large enough. Thus with any finite accuracy of observation the problem is still indeterminate. It is probably this difficulty that has prevented the development of satisfactory methods hitherto.

In practice the observations can give no information about changes of density occurring in a horizontal distance small compared with the depth, and the solution required is one that does not contain such variations. Thus a smoothed solution is sought, the degree of smoothing depending on the accuracy with which it is required to reproduce the observed surface field.

We consider two methods of performing this smoothing, one by the introduction of a smoothing function, the other by a finite difference approximation to Laplace's equation. The processes of computation employed with the two methods are different, and the choice between them depends partly on the facilities available. As explained in § 5, we have found a combination of the two methods to be the most economical.

The analytical method using a smoothing function was devized by G. Kreisel of Trinity College. We originally intended that the present paper should be a joint one with him; it was found, however, that his methods had applications wider than the present problem, and that the mathematical apparatus required for an adequate discussion was best separated from the particular application to gravity anomalies. His work therefore appears as a separate paper (Kreisel 1948).

In § 3 of the present paper a derivation of his expressions is outlined from a point of view that may be more familiar to the physicist than his rigorous discussion. The authors are indebted to him for allowing them to use his work and for his advice and assistance.

3. ANALYTICAL METHOD

(*a*) *Two dimensions.* A simple harmonic gravity variation

$$g_0 = A \cos px$$

can be produced by a distribution of mass of surface density M at a depth z, where,

$$2\pi kM = A\, e^{pz} \cos px. \tag{1}$$

Thus, if gravity at the surface is expressible as a Fourier series, each term can be extrapolated downwards by the use of (1). If the resulting series converges it gives a mass distribution capable of producing the field at the surface. Tsuboi has applied this method to some practical examples (Tsuboi 1937, 1939; Tsuboi & Fuchida 1937, 1938; Tsuboi, Kaneko, Miyamura & Yabasi 1939). It seems unlikely, however, that in this form it can find wide application, as the arithmetic involved in treating a complicated gravity field is altogether prohibitive. What is required is a method analogous to the celebrated deduction by Stokes (1849) of the form of the geoid from the gravity anomalies. The problem is more complicated than Stokes's in that it is necessary to introduce the smoothing discussed in § 2.

Suppose first that the field is constant in a horizontal direction parallel to OY and let x be a co-ordinate measured horizontally in the direction OX in which the field varies, and z a co-ordinate measured vertically downwards. Then

$$g_z(x) = \frac{1}{\pi}\int_0^\infty\int_{-\infty}^\infty g_0(\xi)\, e^{pz}\cos p(\xi - x)\, d\xi\, dp \tag{2}$$

is a solution of Laplace's equation which vanishes when $z \to -\infty$. At $z = 0$ it is the Fourier integral for $g_0(x)$. If the expression (2) exists, it will represent $g_0(x)$ at $z = 0$ and will give a relation between g_0, the field at the surface, and g_z the field at depth z. For the reasons stated in § 2, the integration with respect to p in (2) diverges at the upper limit unless g_0 is a sufficiently smooth function of x.

To avoid this difficulty a smoothed function $\bar{g}_0$ may be substituted for g_0 in (2). A suitable function is

$$\bar{g}_0(\xi) = \sqrt{\frac{\beta}{\pi}}\int_{-\infty}^\infty g_0(x_1)\, e^{-\beta(x_1-\xi)^2}\, dx_1,$$

where β is a constant which determines the severity of the smoothing.

Let the field at a depth z corresponding to this smoothed g_0 be $\bar{g}_z$. Then

$$\bar{g}_z(x) = \frac{\beta^{\frac{1}{2}}}{\pi^{\frac{3}{2}}}\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty g_0(x_1)\, e^{-\beta(x_1-\xi)^2}\, e^{pz}\cos p(\xi - x)\, dx_1\, d\xi\, dp.$$

If the order of integration may be changed, and that with respect to ξ performed first, then

$$\bar{g}_z(x) = \frac{1}{\pi}\int_0^\infty\int_{-\infty}^\infty g_0(x_1)\, e^{-p^2/4\beta}\, e^{pz}\cos p(x_1 - x)\, dx_1\, dp. \tag{3}$$

Taking the p integration next,

$$\bar{g}_z(x) = \int_{-\infty}^\infty g_0(x_1)\, \lambda'(x_1 - x)\, dx_1, \tag{4}$$

where $$\lambda'(x_1 - x) = \frac{1}{\pi}\int_0^\infty e^{-p^2/4\beta}\, e^{pz}\cos p(x_1 - x)\, dp. \tag{5}$$

$\lambda'(x_1 - x)$ is a function of $(x_1 - x)$, z and β which can be tabulated. When this has been done (4) gives $\bar{g}_z(x)$. The surface density required at any depth is then $\bar{g}_z(x)/2\pi k$.

The inversions of the order of integration are easily justified if $\int_{-\infty}^\infty |g_0(x_1)|\, dx_1$ exists and it may be shown that the theorem is true without this restriction provided $g_0(x_1)$ is bounded.

Results identical with (4) and (5) are obtained if, instead of smoothing g_0, we seek a smoothed version of g_z. The necessary changes in the order of the integrations are, however, then more difficult to justify. Kreisel's (1948) paper provides a rigorous treatment from this point of view.

(*b*) *Three dimensions*. The generalization of this argument to three dimensions is straightforward. Take cylindrical co-ordinates r, z, ϕ, where z is measured vertically downwards and the origin is in the surface. In place of (1) consider the expression,

$$e^{pz}J_0(pr)\cos n\phi. \tag{6}$$

If n is an integer this is a solution of Laplace's equation which is finite, tends to zero as $z \to -\infty$ and is single valued.

By analogy with (2) we express $g_z(r)$ as a Bessel-Fourier integral

$$g_z(r,\phi) = \frac{1}{2\pi}\sum_{n=-\infty}^{n=\infty}\int_0^\infty\int_0^{2\pi}\int_0^\infty g_0(r_1,\phi_1)\,e^{pz}J_n(pr_1)\,J_n(pr)\cos n(\phi_1-\phi)\,pr_1\,dr_1\,d\phi_1\,dp. \tag{7}$$

The result is considerably simplified if g_z is calculated at a point below the origin of co-ordinates. This involves no loss of generality as the origin can be placed anywhere. Some simplification is also achieved in the proof by smoothing g_z instead of g_0; though it is then not easy to justify the changes in the order of integration. The smoothed value of g_z is taken as

$$\bar{g}_z(0) = \frac{\gamma}{\pi}\int_0^{2\pi}\int_0^\infty g_z(\rho,\phi_2)\,e^{-\gamma\rho^2}\rho\,d\rho\,d\phi_2,$$

where γ is a smoothing parameter analogous to β. Substituting in (7) and integrating with respect to ϕ_2 removes all terms except those with $n = 0$. The ϕ_1 integration averages g_0 round a circle and

$$\bar{g}_z(0) = 2\gamma\int_0^\infty\int_0^\infty\int_0^\infty g_0(r_1)\,e^{pz}J_0(pr_1)\,J_0(p\rho)\,e^{-\gamma\rho^2}\rho p r_1\,dr_1\,dp\,d\rho,$$

where $$g_0(r_1) = \frac{1}{2\pi}\int_0^{2\pi} g_0(r_1,\phi_1)\,d\phi_1.$$

The ρ integration can now be performed giving,

$$\bar{g}_z(0) = \int_0^\infty g_0(r_1)\,\Lambda'(r_1)\,dr_1, \tag{8}$$

where $$\Lambda'(r_1) = r_1\int_0^\infty e^{pz}\,e^{-p^2/4\gamma}J_0(pr_1)\,p\,dp. \tag{9}$$

Here Λ' is a function of r, z, γ, which is closely analogous to λ' in (5).

(*c*) *Solution for a sphere*. The solution can be obtained for a spherical surface by a closely similar argument starting from

$$\left(\frac{R_1}{R}\right)^{p+2} P_p^m(\cos\theta)\cos m\phi,$$

instead of (1) or (6), where θ and ϕ are spherical polar co-ordinates relative to an axis through the point at which the mass is to be found; p and m are integers, R is the radius of the sphere over which g_0 is measured and R_1 is that of the sphere over which the mass distribution is sought. The smoothed value of g_z is taken as

$$\bar{g}_z(0) = \frac{\gamma e^{2\gamma}}{2\pi \sinh(2\gamma)} \int_0^{\infty} \int_0^{2\pi} g_z(\theta_2, \phi_2)\, e^{-4\gamma \sin^2 \frac{1}{2}\theta_2} \sin\theta_2\, d\phi_2\, d\theta_2,$$

where γ is a suitable smoothing parameter. The result found is then

$$\bar{g}_z(0) = \int_0^{\pi} g_0(\theta_1)\, \Lambda_1(\theta_1)\, d\theta_1,$$

where $g_0(\theta_1)$ is the average value of $g_0(\theta, \phi)$ round a circle of angular radius θ_1, and

$$\Lambda_1 = \frac{\sqrt{(\pi\gamma)}\sin\theta_1}{2\sinh(2\gamma)} \sum_{p=0}^{\infty} (2p+1) \left(\frac{R}{R_1}\right)^{p+2} P_p(\cos\theta_1)\, I_{p+\frac{1}{2}}(2\gamma).$$

The corresponding mass distribution M_z is no longer $g_z/2\pi k$, but is given by

$$\bar{M}_z(0) = \int_0^{\pi} g_0(\theta_1)\, \Lambda_2(\theta_1)\, d\theta_1,$$

where $$\Lambda_2(\theta_1) = \frac{\sqrt{\gamma}\sin\theta_1}{8\sqrt{\pi}\, k \sinh(2\gamma)} \sum_{p=0}^{\infty} \frac{(2p+1)^2}{p+1} \left(\frac{R}{R_1}\right)^{p+2} P_p(\cos\theta_1)\, I_{p+\frac{1}{2}}(2\gamma).$$

$I_{p+\frac{1}{2}}$ is the Bessel function of imaginary argument and half integer order. Λ_1 and Λ_2 may be shown to reduce to (9) when g_0 is zero, except over an area whose dimensions are small compared with the radius of the sphere.

4. TABULATION OF λ AND Λ

Before tabulating λ' and Λ' it is convenient to render them non-dimensional. The depth z at which (5) or (9) is evaluated forms the natural unit of length in terms of which the other quantities may be expressed. For a reasonable degree of smoothing $z\sqrt{\beta}$ and $z\sqrt{\gamma}$ will be numbers of the order of unity; call these $\sqrt{\sigma}$ and $\sqrt{\tau}$.

If x and x_1 are expressed in units of z, (4) and (8) will still hold if,

$$\lambda(x_1 - x) = \frac{1}{\pi} \int_0^{\infty} e^{p}\, e^{-p^2/4\sigma} \cos p(x_1 - x)\, dp, \tag{10}$$

and $$\Lambda(r_1) = r_1 \int_0^{\infty} e^{p}\, e^{-p^2/4\tau} J_0(pr_1)\, p\, dp \tag{11}$$

are substitued for λ' and Λ'; the depth now no longer occurs explicitly.

The function λ can be connected with one tabulated by Born (1933, p. 486), by Mitchell & Zemansky (1934, p. 329), and by Hjerting (1938), in investigations of optical absorption. Born tabulates

$$\psi(x, \eta) = \frac{1}{\eta\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{1}{1+y^2} \exp\left[-\left(\frac{x-y}{\eta}\right)^2\right] dy$$

which is connected with λ by

$$\lambda(x) = 2\sqrt{(\sigma/\pi)} \cos 2\sigma x \exp[\sigma(1 - x^2)] - (1/\pi)\, \psi(x, 1/\sqrt{\sigma}).$$

Professor Born informs us that only the first two figures of his values of ψ are reliable. We have therefore only used this expression for a preliminary study of the function λ. The values for $\sigma = 1$, which are those used for the practical application of the method, have been calculated *ab initio* by expansion in powers of x and $1/x$ and by numerical integration. The results* are given in table 1 and figure 1.

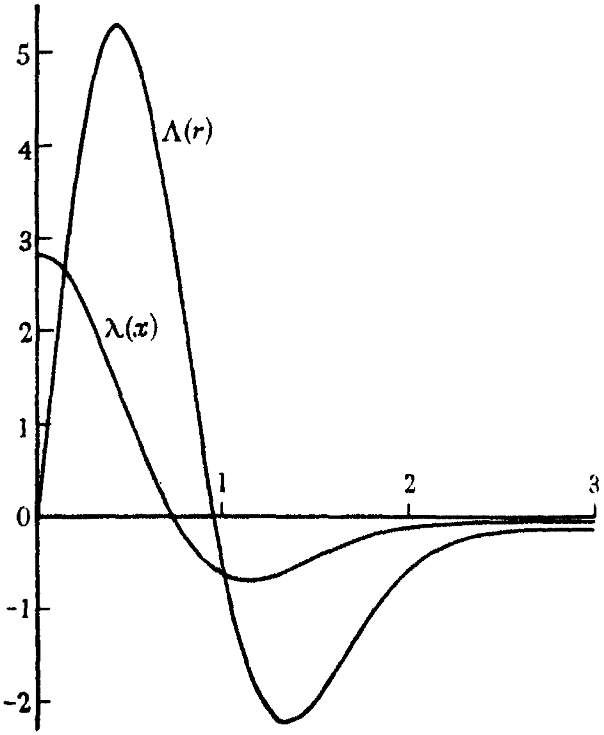


FIGURE 1. $\lambda(x)$ for $\sigma = 1$ and $\Lambda(r)$ for $\tau = 1$.

The function Λ has been computed for $\tau = 1$ by numerical integration and by expansion in powers of r and $1/r$. The results are given in table 1 and figure 1.

Beyond the limits of table 1 the following expressions will give results accurate to four places of decimals:

$$\pi\lambda(x) \sim -1/x^2 - (3/2\sigma - 1)/x^4,$$

$$\pi\int_0^x \lambda(x_1)\,dx_1 \sim \tfrac{1}{2}\pi + 1/x + (3/2\sigma - 1)/3x^3,$$

$$\Lambda(r) \sim -1/r^2 - 3(3/2\tau - 1)/2r^4,$$

$$\int_0^r \Lambda(r_1)\,dr_1 \sim 1 + 1/r + (3/2\tau - 1)/2r^3.$$

The columns headed M'' in table 1 give modified second differences (Comrie 1936, p. 928) with the aid of which the table may be interpolated. As third differences are

---

* The computation of most of the results in table 1 has been carried out by Miss Mumford and Miss Munns of the Mathematical Laboratory, to whom our thanks are due. Mr Jones kindly computed the coefficients of the ascending series for $\Lambda(r)$ up to the term in r^{11}.

TABLE 1

x or r	$\lambda(x)$ $\sigma = 1$	M''	$\int_0^x \lambda(x_1)\,dx_1$ $\sigma = 1$	M''	$\lambda(x)$ $\sigma = 4$	$\Lambda(r)$ $\tau = 1$	M''	$\int_0^r \Lambda(r_1)\,dr_1$ $\tau = 1$	M''
0·0	2·8260	−1848	0·0000	0	122·925	0·0000	0	0·0000	2006
0·1	2·7358	−1728	0·2796	−181	82·191	1·9277	−2865	0·0976	1860
0·2	2·4766	−1402	0·5415	−338	−3·349	3·5789	−5190	0·3763	1449
0·3	2·0801	−917	0·7703	−455	−63·665	4·7289	−6550	0·7965	875
0·4	1·5932	−361	0·9545	−520	−65·125	5·2452	−6755	1·3008	177
0·5	1·0699	180	1·0877	−528	−29·882	5·1062	−5866	1·8236	−459
0·6	0·5629	631	1·1690	−487	−3·065	4·3958	−4151	2·3029	−968
0·7	0·1162	930	1·2023	−408	13·240	3·2778	−2030	2·6891	−1274
0·8	−0·2404	1073	1·1953	−306	9·259	1·9565	76	2·9517	−1366
0·9	−0·4926	1057	1·1578	−200	2·752	0·6355	1785	3·0805	−1275
1·0	−0·6413	921	1·1002	−98	−0·495	−0·5182	2890	3·0844	−1034
1·1	−0·6992	715	1·0325	−16	−0·942	−1·3953	3320	2·9862	−723
1·2	−0·6861	482	0·9628	41	−0·520	−1·9511	3170	2·8161	−388
1·3	−0·6246	260	0·8969	81	−0·205	−2·1977	2595	2·6063	−102
1·4	−0·5364	78	0·8387	97	−0·104	−2·1886	1817	2·3851	123
1·5	−0·4395	−53	0·7899	98	−0·091	−1·9983	1011	2·1746	261
1·6	−0·3469	−128	0·7507	86	−0·091	−1·7047	326	1·9889	326
1·7	−0·2663	−161	0·7201	75	−0·087	−1·3750	−168	1·8348	335
1·8	−0·2012	−160	0·6969	57	−0·080	−1·0586	−474	1·7134	300
1·9	−0·1517	−138	0·6794	41	−0·073	−0·7862	−592	1·6216	246
2·0	−0·1159	−112	0·6661	29	−0·068	−0·5706	−578	1·5543	185
2·1	−0·0912	−82	0·6558	21	−0·062	−0·4116	−499	1·5056	133
2·2	−0·0746	−55	0·6476	12	−0·057	−0·3018	−381	1·4703	88
2·3	−0·0636	−36	0·6407	9	−0·053	−0·2302	−271	1·4440	55
2·4	−0·0563	−22	0·6347	7	−0·050	−0·1859	−171	1·4234	32
2·5	−0·0512	−13	0·6294	4	−0·046	−0·1593	−103	1·4062	22
2·6	−0·0474	−5	0·6245	—	−0·043	−0·1434	−53	1·3912	11
2·7	−0·0442	−4	0·6199	—	−0·040	−0·1332	−24	1·3774	7
2·8	−0·0414	—	0·6156	—	−0·037	−0·1257	−10	1·3644	8
2·9	−0·0389	—	0·6116	—	−0·035	−0·1193	−1	1·3522	6
3·0	−0·0365	—	0·6078	—	−0·033	−0·1132	—	1·3406	5
3·1	−0·0343	—	0·6043	—	−0·031	−0·1072	—	1·3295	7
3·2	−0·0322	—	0·6010	—	−0·029	−0·1014	—	1·3191	5
3·3	−0·0303	—	0·5978	—	−0·028	−0·0959	—	1·3092	6
3·4	−0·0285	—	0·5949	—	−0·026	−0·0906	—	1·2999	5
3·5	−0·0269	—	0·5921	—	−0·025	−0·0856	—	1·2911	5
3·6	−0·0254	—	0·5895	—	−0·024	−0·0809	—	1·2828	4
3·7	−0·0240	—	0·5870	—	−0·022	−0·0765	—	1·2749	5
3·8	−0·0228	—	0·5847	—	−0·021	−0·0724	—	1·2675	3
3·9	−0·0216	—	0·5825	—	−0·020	−0·0686	—	1·2604	4
4·0	−0·0205	—	0·5804	—	−0·019	−0·0651	—	1·2537	4
4·1	−0·0195	—	0·5784	—	−0·018	−0·0619	—	1·2474	—
4·2	−0·0185	—	0·5765	—	−0·017	−0·0589	—	1·2413	—
4·3	−0·0177	—	0·5747	—	−0·017	−0·0561	—	1·2356	—
4·4	−0·0168	—	0·5730	—	−0·016	−0·0535	—	1·2301	—
4·5	−0·0161	—	0·5713	—	−0·015	−0·0511	—	1·2249	—
4·6	−0·0154	—	0·5697	—	−0·015	−0·0488	—	1·2199	—
4·7	−0·0147	—	0·5682	—	−0·014	−0·0467	—	1·2151	—
4·8	−0·0141	—	0·5668	—	−0·013	−0·0447	—	1·2105	—
4·9	−0·0135	—	0·5654	—	−0·013	−0·0429	—	1·2062	—
5·0	−0·0130	—	0·5641	—	−0·012	−0·0411	—	1·2020	—

not negligible Everett's formula should be used. The column giving $\lambda(x)$ for $\sigma = 4$ cannot be interpolated.

Table 1 shows that for large values of σ, that is for slight smoothing, λ has a large maximum for small values of x followed by a deep minimum; Λ behaves in a similar way. Thus an irregularity in g_0 that is of extent small compared with the depth will yield a large positive mass surrounded by large negative masses. This is a mild manifestation of the indeterminacy discussed in § 2. It is found to be impracticable to use a value of σ or τ much greater than 1, and the most convenient method of computation is to proceed downward in steps using σ or $\tau = 1$ for each step. If at any stage of the calculation some feature of the field becomes improbably prominent it can be left at a shallow depth and the rest of the field carried downward. The appearance of large neighbouring anomalies of opposite sign is an indication that an attempt is being made to find a solution at too great a depth. In nature a field of anomalies will usually be due to both shallow and deep masses, and the stepwise procedure is necessary in order that the shallow parts may be left at their proper depths. The process is necessarily to some extent arbitrary, for a shallow limit can never be set to the depth at which the masses lie from a study of gravity alone.

The integrals (4) and (8) may most readily be evaluated by breaking them up into short ranges over which g_0 may be put equal to its mean value over the range or to its value at the centre of the range,

$$\bar{g}_z(0) = \Sigma g_0(x_{s+\frac{1}{2}}) \int_{x_s}^{x_{s+1}} \lambda(x_1)\,dx_1 \quad \text{or} \quad \Sigma g_0(r_{s+\frac{1}{2}}) \int_{r_s}^{r_{s+1}} \Lambda(r_1)\,dr_1.$$

$\int_0^x \lambda(x_1)\,dx_1$ and $\int_0^r \Lambda(r_1)\,dr_1$ are given in table 1 and a suitable set of intervals for the first integral in table 2. For three-dimensional problems it is convenient to arrange the intervals of the r_1 integration so that the integrals take simple fractional values. Suitable limits are given in table 3. The field g_z may then be calculated by

TABLE 2

x_1	x_2	$\int_{x_1}^{x_2} \lambda(x)\,dx$	x_1	x_2	$\int_{x_1}^{x_2} \lambda(x)\,dx$
0·0	0·25	0·661	3	4	−0·028
0·25	0·5	0·427	4	5	−0·016
0·5	1·0	0·012	5	10	−0·032
1·0	1·5	−0·310	10	20	−0·016
1·5	2·0	−0·124	20	∞	−0·016
2	3	−0·058			

TABLE 3

r_1	r_2	$\int_{r_1}^{r_2} \Lambda(r)\,dr$	r_1	r_2	$\int_{r_1}^{r_2} \Lambda(r)\,dr$
0·00	0·23	0·5	1·35	1·59	−0·5
0·23	0·34	0·5	1·59	2·12	−0·5
0·34	0·44	0·5	2·12	2·54	−0·1
0·44	0·54	0·5	2·54	3·40	−0·1
0·54	0·65	0·5	3·40	5·05	−0·1
0·65	0·83	0·5	5·05	10·02	−0·1
0·83	1·35	−0·5	10·02	∞	−0·1

making a template showing the limits given in table 3 and superposing this on a map of the anomalies. The mean anomalies in the successive zones are then estimated, multiplied by the factors from table 3 and added.

To obtain the corresponding masses in feet or metres of rock of density D the values of g_s are multiplied by $24/D$ m./milligal or $78/D$ ft./milligal.

5. Finite difference method

(*a*) *Two dimensions.* An alternative method of solving the problem is to use a finite difference approximation to Laplace's equation and to integrate downward by a numerical process.

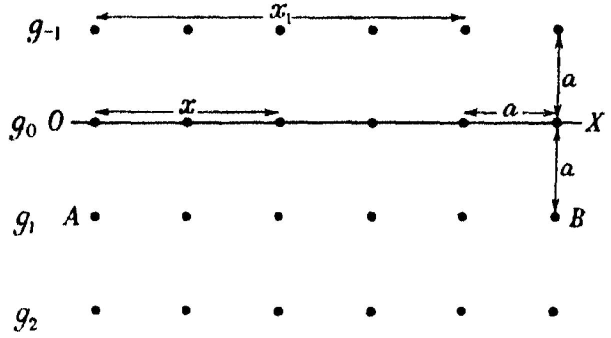


Figure 2

Suppose the vertical field $g_0(x)$ given at a set of points spaced at equal intervals a along a horizontal line OX (figure 2). Let the field be constant in a horizontal direction at right angles to this line. The field $g_{-1}(x_1)$ at a point on a line at a distance a above OX, is

$$g_{-1}(x_1) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{a g_0(x)\,dx}{a^2+(x_1-x)^2}. \tag{12}$$

If a is chosen small enough, g_0 will differ little between neighbouring points and may be assumed constant and equal to $g_0(na)$ over the range $(n \pm \frac{1}{2})a$. Then

$$g_{-1}(sa) = \frac{1}{\pi}\sum_{n=-\infty}^{n=\infty} g_0(na)\{\tan^{-1}(n-s+\tfrac{1}{2})-\tan^{-1}(n-s-\tfrac{1}{2})\}. \tag{13}$$

The field is now known on two lines of points forming part of a square grid. Let g_1 be the field at the points on the line AB of this grid next below OX. If the masses producing the field lie below this line, the field can be extrapolated downward by using a finite difference approximation to Laplace's equation. The simplest is to take g at each point to be the mean of the values at its four nearest neighbours in the lattice (Southwell 1946),

$$g_1(sa) \doteq 4g_0(sa) - g_0(sa-a) - g_0(sa+a) - g_{-1}(sa). \tag{14}$$

The surface density along the line AB that will produce the observed field along OX is then $g_1/2\pi k$.

The values of g on the second and lower lines of the lattice can be found by a repetition of this process, giving a series of mass distributions at depths a, $2a$, ..., each capable of producing the observed field.

It is evident that in this method no variations in the gravity field occurring within a distance of less than a mesh-length can affect the calculated mass-distributions. In this way very rapid variations in g_z are prevented as effectively as in the analytical method.

In (14) g_0 is multiplied by a factor 4, and this occurs again at each step in the downward integration. This leads to a rapid magnification of small irregularities in the initial data and to an instability similar to that discussed in § 4. The process may be kept in control by choosing a so that only 3 or 4 steps are needed to reach the required depth and by smoothing at each step. The smoothing is best carried out by a process not involving judgement; forming a running mean of three successive values, or substracting $\frac{1}{12}$ of the fourth differences (Jeffreys 1934; Comrie & Jeffreys 1932), are suitable processes and two successive applications of them at each depth are usually sufficient. As with the analytical method it is often necessary to leave certain features at shallow depths.

The downward steps by the use of (14) are much more rapid than the corresponding steps in the analytical method using (4). The first upward step by (13) is, however, as tedious as one downward step by (4). There is thus some economy in time by using the analytical method to make one downward step and then proceeding downwards by the finite difference method. This is the method that we have used in practical applications.

(*b*) *Three dimensions.* The extension of the method to three dimensions is simple in principle. All that is necessary is to substitute a surface integral for (12) and to make each g equal to the mean of the values at the six neighbouring points. We have not used this method in practice.

6. MAXIMUM DEPTH FOR MASSES

There is a limit to the depth that can be assumed for the material which produces a disturbance of gravity. This enables a limit to be set to the depth at which the disturbing masses can lie. Consider, for example, a horizontal sheet of small thickness h lying at a depth z and having a density $\rho \cos px$ which varies in a horizontal direction. The disturbance of gravity at the surface is then $g \cos px$, where

$$g = 2\pi k\rho h e^{-pz}.$$

If ρ_m is the maximum value that can plausibly be ascribed to ρ then

$$z < p^{-1} \log 2\pi k\rho_m h/g.$$

If the value of gravity at the surface does not vary simple harmonically an analogous inequality will apply to the components into which the field can be analyzed by forming the Fourier transform. A discussion from this point of view is given by Kreisel (1948, theorem B). The rigorous application of his result is complicated, but for practical purposes it usually suffices to choose the greatest gravity gradient for which we wish to account. Suppose that between the maximum and the minimum lying on each side of this maximum gradient the anomaly changes by g_m in a distance x_m. Then the Fourier analysis of the field will contain terms with wave-lengths at least as short as $2x_m$ (i.e. $p = \pi/x_m$). The amplitude in the neighbourhood of this wave-length will be about $\frac{1}{2}g_m$. The upper limit to the depth, therefore, becomes

$$z < \pi^{-1}x_m \log 4\pi k\rho_m h/g_m.$$

If, instead of a sheet of thickness h, we assume a body of indefinitely great thickness with its upper surface at a depth z, then a similar argument gives

$$z < \pi^{-1}x_m \log 4k\rho_m x_m/g_m.$$

This expression is applied in § 8.

7. ARTIFICIAL EXAMPLES

The behaviour of the methods can be illustrated by applying them to artificial examples in which g_0 is assumed to be the field produced by some simple distribution of mass. The values of $\bar{g}_z$ are then compared with the true field calculated from the mass distribution.

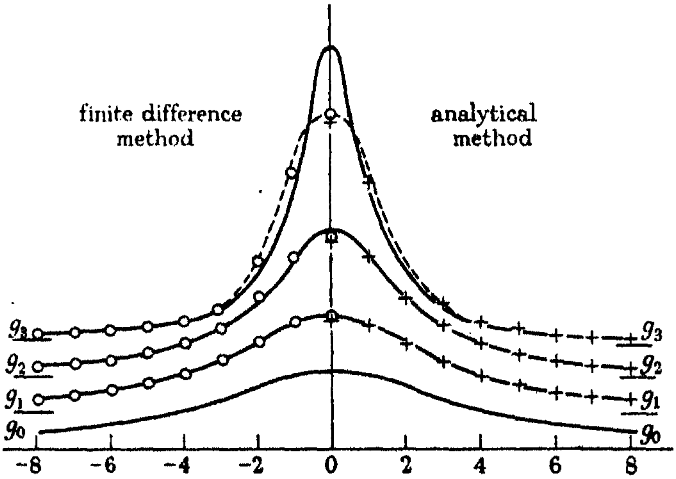


FIGURE 3. Comparison of the results obtained by the analytical and finite difference methods with the calculated field due to a horizontal line mass at a depth of four units. —— exact result, + analytical method, ⊙ finite difference method.

Figures 3 and 4 show the results for a line mass. The left-hand sides of the figures show the results obtained by the finite difference method, using a mesh with a side equal to a quarter of the depth of the mass. The right-hand sides show those from the analytical method with $1/\sqrt{\beta}$ equal to the depth of the mass. The results of the two methods are closely similar; g_1 and g_2 are good approximations to the field at depths of $\frac{1}{4}$ and $\frac{1}{2}$ that of the mass, their maximum errors being 1·2 and 6 % for the finite difference method and 6 and 9 % for the analytical method. For g_3 the maximum errors are 25 %; as the minimum distance from the line mass is now only equal to the length of the side of the mesh used in the finite difference method a large error is to be expected. To illustrate the behaviour of the solution at depths equal to, and greater than, that of the line mass, the calculations have been continued to give g_4 and g_5 (figure 4). The former shows a single peak that may be regarded as obtained by smoothing the line mass. The latter shows a large peak below the line mass bordered by narrow negative strips on each side. If the calculation were continued to greater depths it would show still more violent oscillations. This example suggests that the occurrence of alternating positive and negative signs in the calculated field is a sensitive indication of the maximum possible depth for the masses producing a given anomaly.

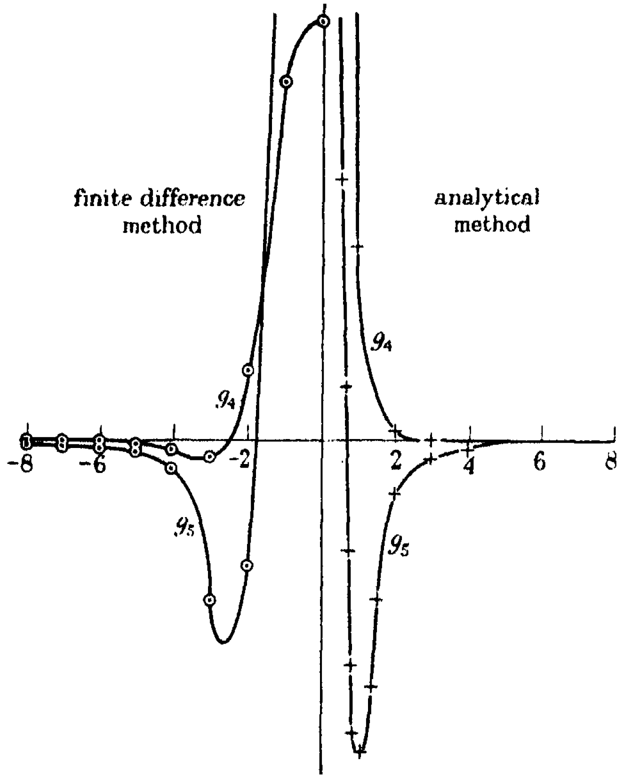


FIGURE 4. Results obtained by the analytical and finite difference methods from the field produced by a line mass at a depth of four units, showing a large oscillation in g_5. + analytical method, ⊙ finite difference method.

A line mass is an unfavourable example for the method and is also not a type of field that occurs in practice. It is therefore of interest to consider an example in which the mass is less concentrated. For this we take a semi-infinite sheet with its edge at a depth of four times the length of the side of the square mesh used in the finite difference method. The calculation has been made exactly as in the real examples of § 8: that is the analytical method with $\sigma = 1$ has been used for one step and the finite difference method for the remaining steps. The results for $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$ of the depth of the mass (figure 5) are excellent approximations to the calculated curves, except in the immediate neighbourhood of the edge of the mass. The calculated mass distribution at the depth of the sheet, g_4, smooths out the sudden transition near the edge, but gives good agreement at all distances beyond two mesh lengths from the edge. The calculated points show some scatter due to the accentuation of rounding off errors in the calculation.

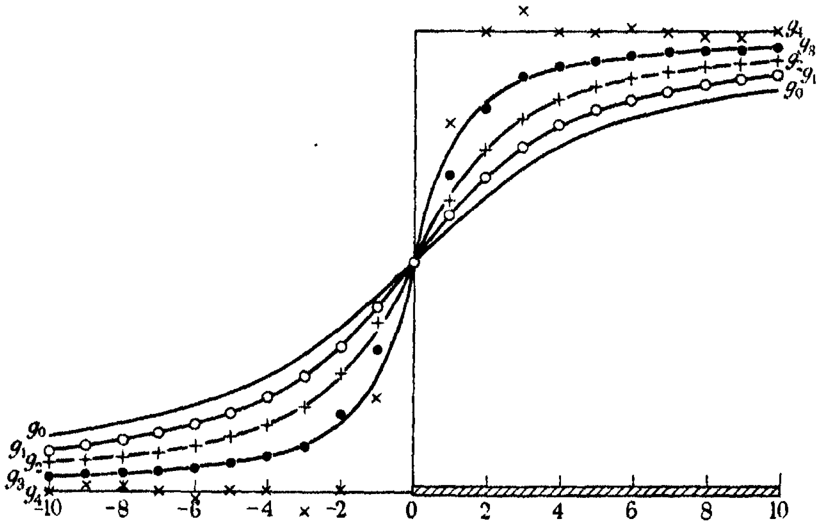


FIGURE 5. Comparison of the results obtained with the calculated field due to a horizontal semi-infinite slab. ○ $= g_1$, calculated by the analytical method from g_0; + $= g_2$, calculated by the finite difference method from g_0 and g_1; ● $= g_3$, calculated by the finite difference method from g_1 and g_2; × $= g_4$, calculated by the finite difference method from g_2 and g_3.

8. APPLICATION TO THE ALPS

As an example of the application of the method to a real problem, we have studied a cross-section of the Alps from Lugano to the Black Forest (figures 6 and 7). The lines of equal anomaly are taken from a paper by Niethammer (1921),* and the section prepared from them. The section is at right angles both to the strike of the main Alpine structures and to the lines of equal anomaly. Except at its south-eastern end it may, to a good approximation, be regarded as providing a two-dimensional problem. The topographic section shown in figure 7 is constructed from the map of average heights in 8 km. squares given by Niethammer & Lalive (1925).

As a first step in the downward integration the analytical method was used with $\sigma = 1$ to find the field at a depth of 20 km. The result, which is shown in figure 7, is similar to the original curve but steeper and has a pronounced peak in the neighbourhood of Lugano. Examination of figure 6 shows that the small hump in the surface field in this region is at the north-east end of a positive strip becoming more conspicuous to the south-west. A section taken farther to the south-west would have shown an even greater accentuation of the bulge in the 20 km. curve: there is, therefore, little doubt that this feature is of shallow origin.

---

* 10 milligals has been subtracted from all the anolamies to allow for the error in the base station used by Niethammer (see Heiskanen 1939) and a further 14 milligals to convert to the international gravity formula.

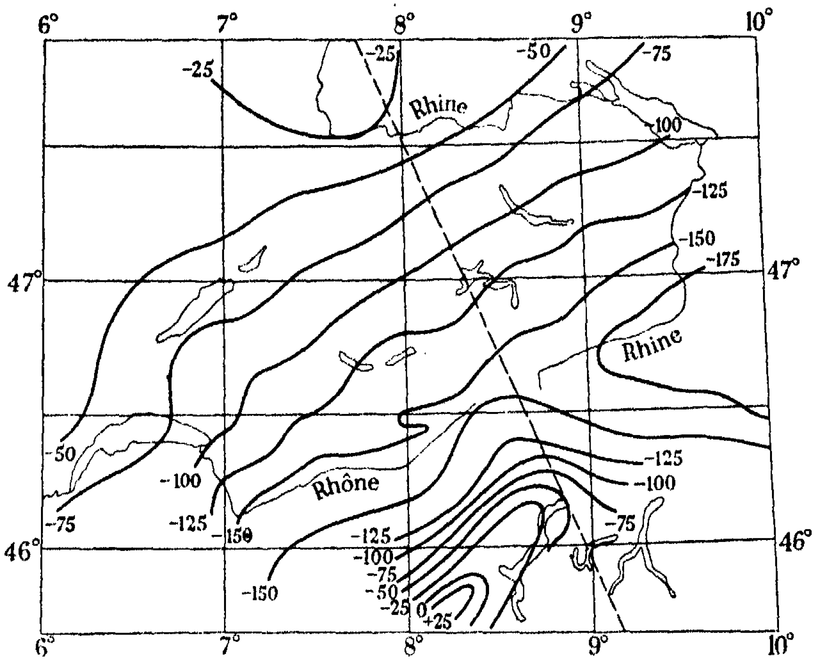


Figure 6. Bouguer anomalies in Switzerland (in milligals).
– – – – = line of section for figure 7.

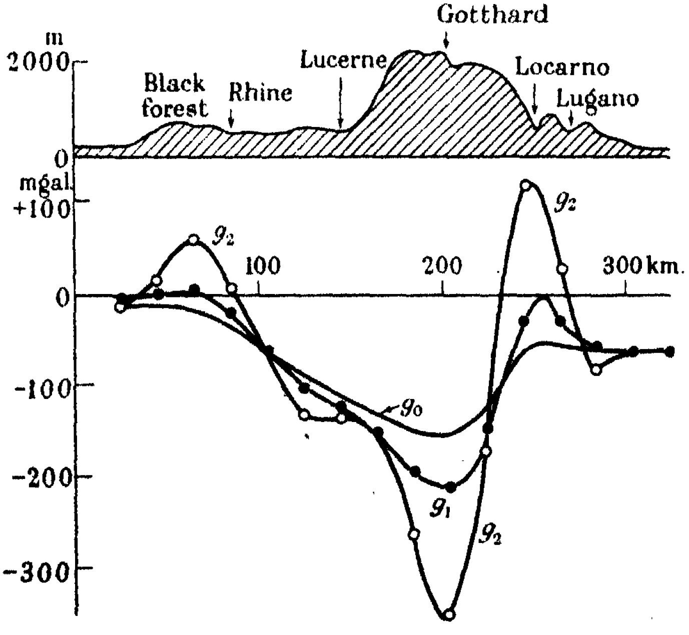


Figure 7. Topographic section and gravity anomalies along the line shown in figure 6.
g_1 = field at 20 km. depth: g_2 = field at 40 km. depth.

A second step from 20 to 40 km. was taken by the finite difference method and gave the second curve shown. The hump near Lugano now extends to about + 150 milligals and the neighbouring trough to − 350 milligals, a change of 500 milligals in 45 km. The maximum density difference that can plausibly be assumed to account for the anomalies is about 0·3 g./cm.^{3}. With this difference a change of 500 milligals would require 40 km. of rock. Farther to the south-west the requirements would be still more extreme. On the north side of the Gotthard also, the curve has become improbably steep.

The method of § 6 may be used to set a limit to the depth of the masses causing the rapid change of anomaly near the south-east end of the line. In the 60 km. between the minimum near the Gotthard to the maximum near Locarno the Bouguer anomaly changes by 110 milligals. If the maximum density difference is 0·3 g./cm.^{3} the maximum depth given by (12) is 28 km.

A comparatively shallow origin for the anomalies agrees with recent gravitational studies in other areas and with the seismological evidence on the thickness of the crustal layers. The gravity field and the mass distributions calculated from it show a well-marked asymmetry with steep gradients on the south-east and a gradual rise on the north-west. The negative anomalies are displaced to the north of the central axis of the mountains and extend under the Swiss plain. The south-east margin of the mountains appears to be affected by an excess of mass lying near the 'zone of roots' of the Pennine nappes. Such an asymmetrical arrangement is accordant with the known structure of the mountains, and suggests that the northward thrusts by which they are known to have been formed have brought dense basic rocks near the surface to the south-east, and that the weight of the overthrust mountains has bowed down the crust to the north-west causing the isostatic compensation to extend beyond them and leaving a basin for the deposition of the molasse.* The mass excess to the south-east may be compared to that occurring on the inside of the East Indian arc, and perhaps to the positive anomalies associated with the Troodos mountains of Cyprus.

It is not possible in this paper to enter into a more detailed discussion of the relation of the gravity field of Switzerland to Alpine structure, though it would seem well worthy of fresh consideration. The existing discussions (Lehner 1929) are somewhat severely geodetic.

9. CONCLUSIONS

The investigations described in §§ 7 and 8, and others that we hope to publish elsewhere, suggest that the methods developed in the present paper are practicable and may frequently be useful. It may be desirable, however, to emphasize that they do not always repay the considerable arithmetical work involved. If a gravity anomaly is due to masses that are shallow, relative to the least horizontal distance in which the anomaly changes by a large fraction of its whole amplitude, then a mass distribution $g/2\pi k$ is an excellent and simple approximation. At the other extreme, a depth large compared with the horizontal distance in which the anomaly changes appreciably leads to an impossible distribution of masses. The examination of specific fields of anomalies by the methods of this paper shows that the gap between these two extremes is usually a narrow one and that the main use of the method is in setting a limit to the possible depth of the masses.

---

* Some of the negative anomaly under the plain may be due to the lightness of the molasse which has an average density of about 2·5.

REFERENCES

Born, M. 1933 *Optik*. Berlin: Julius Springer Verlag.

Cassinis, G., Dore, P. & Ballarin, S. 1937 *R. Comm. Geod. Ital. Pub.* N.S. no. 13.

Comrie, L. J. 1937 *Interpolation and allied tables*. London: H.M. Stationery Office.

Comrie, L. J. & Jeffreys, H. 1932 *Mon. Not. R. Astr. Soc. Geophys. Suppl.* **3**, 201–212.

Heiskanen, W. 1939 *Ann. Acad. Sci. fenn.* A, **51**, no. 10.

Hjerting, F. 1938 *Astrophys. J.* **88**, 508–515.

Jeffreys, H. 1934 *Proc. Camb. Phil. Soc.* **30**, 134–138.

Jung, K. 1927 *Z. Geophys.* **3**, 201–212.

Kreisel, G. 1948 *Proc. Roy. Soc.* A (in the Press).

Lehner, M. 1929 *Verh. naturf. Ges. Basel*, **61**. 1–76.

Meinesz, F. A. Vening. 1934 *Gravity expeditions at sea*, vol. 2, pp. 23–28. Delft: J. Waltman, Jr.

Mitchell, A. C. G. & Zemansky, M. W. 1934 *Resonance radiation and excited atoms*. Cambridge University Press.

Niethammer, Th. 1921 *Astr.-geod. Arb. Schweiz*, p. 16.

Niethammer, Th. & Lalive, A. 1925 *Astr.-geod. Arb. Schweiz*, p. 17.

Numerov, B. 1927 *Bull. Inst. astr. Leningr.* **15**, 165–167.

Southwell, R. V. 1946 *Relaxation methods in theoretical physics*. Oxford: Clarendon Press.

Stokes, C. G. 1849 *Trans. Camb. Phil. Soc.* **8**, 672. (*Math. and Phys. Papers*, **2**, 131–171.)

Tsuboi, C. 1937 *Bull. Earthq. Res. Inst. Tokyo*, **15**, 650–653.

Tsuboi, C. 1939 *Bull. Earthq. Res. Inst. Tokyo*, **17**, 351–384.

Tsuboi, C. & Fuchida, T. 1937 *Bull. Earthq. Res. Inst. Tokyo*, **15**, 636–649.

Tsuboi, C. & Fuchida, T. 1938 *Bull. Earthq. Res. Inst. Tokyo*, **16**, 273–284.

Tsuboi, C., Kaneko, T., Miyamura, S. & Yabasi, T. 1939 *Bull. Earthq. Res. Inst. Tokyo*, **17**, 385–410.

The velocity of propagation of electromagnetic waves derived from the resonant frequencies of a cylindrical cavity resonator

By L. Essen, D.Sc., Ph.D. and A. C. Gordon-Smith
The National Physical Laboratory

(*Communicated by Sir Charles Darwin, F.R.S.—Received* 4 *December* 1947)

The frequency of resonance of an evacuated cavity resonator in the form of a right circular cylinder is given by the formula

$$f = v_0\sqrt{\left[\left(\frac{r}{\pi D}\right)^2+\left(\frac{n}{2L}\right)^2\right]}\left[1-\frac{1}{2Q}\right],$$

in which v_0 is the free-space velocity of electromagnetic waves, D and L are the internal diameter and length respectively of the cylinder, r is a constant for a particular mode of resonance, n is the number of half-wave-lengths in the resonator and Q is the quality factor. Assuming the validity of this equation the value of v_0 can be obtained from measured values of f, D, L and Q.

A copper cylinder of diameter approximately 7·4 cm. and length 8·5 cm. was constructed with the greatest uniformity of diameter and squareness of end-faces and its dimensions were measured. The resonant frequencies for a number of different modes were measured and experiments were made to show that the effects on frequency of the coupling probes to the oscillator and detector were negligibly small. It was concluded from these measurements that the most favourable experimental conditions can be obtained for the E_{010} and E_{011} modes. Final measurements on these gave

$$v_0 = 299{,}792 \text{ km./sec.}$$

The estimated maximum error of the result is 9 km./sec. (3 parts in 10^5). This is the error of a single measurement and, since most of the errors are not necessarily random, little is gained by making a large number of measurements. The value is 16 km./sec. greater than the recently determined values of the velocity of light, although the results are not in disagreement when the combined limits of accuracy are taken into account.

1. INTRODUCTION

The velocity of propagation of electromagnetic waves is, in Maxwell's theory, given by $1/\sqrt{(\mu\epsilon)}$, μ being the permeability and ϵ the permittivity of the medium in rationalized practical units. The value for a vacuum, $1/\sqrt{(\mu_0\epsilon_0)}$, is a constant of fundamental importance, being theoretically identical with the velocity of light, and appearing in a great number of electrical calculations. During recent years its value has also become of significance in the field of radio engineering, since it is required for the calculation of the resonant frequencies of electrical circuits, such as cavity resonators, and also for methods of navigation depending on radio waves. The measurement of this constant has attracted many research workers, and possibly more effort has been devoted to it than to the determination of any other general constant. In most of the experiments that have been performed the velocity of light has been determined directly by the measurement of the time occupied in its travel over a certain distance. In all cases the time interval is produced by a revolving or alternating mechanism so that the quantities actually measured are frequency and distance. Both of these can be measured with a high accuracy—better than 1 part in 10^5—and the main experimental difficulty appears to be the adjustment of the variable frequency or length to correspond to the fixed length or frequency as the case may be. In order to eliminate as far as possible the random observational errors a great number of determinations are necessary. Even so it was found in the two most recent experiments (Michelson, Pease & Pearson 1935; Anderson 1941) that the means of groups of a considerable number of measurements differed by amounts greater than would be expected if the errors were entirely random. Michelson *et al.* found differences of 11 km./sec. between the means of groups consisting of about 500 measurements and a maximum difference of 93 km./sec. between the means of groups of about six measurements. Anderson does not give the individual results, but he obtained a difference of 67 km./sec. between the means of groups of about 200 measurements. The close agreement between their final results of 299,774 ± 11 and 299,776 ± 14 km./sec. may therefore be misleading.

The constant has also been determined by a comparison between the calculated and measured values of the capacitance of a capacitor of suitable shape (Rosa & Dorsey 1907). The value calculated from electromagnetic theory involves the permittivity of the medium and the dimensions of the capacitor. The measured value, obtained by balancing a bridge network of which the capacitor forms one arm, is expressed in terms of the international ohm. The experiment consists in the measurement of the dimensions of the capacitor, and of a resistance, a resistance ratio, and a frequency, and the result can be interpreted as the measured value of the permittivity of free space. From this the velocity can be calculated, the value of permeability being known by definition. Both the calculations and experiments are beset with difficulties, but the result obtained by Rosa & Dorsey was probably the most reliable up to that time. The final value given is 299,710 km./sec., which is the mean of about 900 individual determinations, the average deviation is 22 km./sec. and the estimated maximum error is 30 km./sec., apart from uncertainties in the value taken for the international ohm. This latter constant is now known with an accuracy of about 2 parts in 10^5, and Birge (1941) has applied a correction to their results to obtain a value of 299,784 km./sec. Using all the reliable data existing at the time Birge gives the most probable value for the constant as

$$c = 299{,}776 \pm 4\,\text{km./sec.}$$

In recent years a number of experimenters have measured the speed of travel of electric waves of frequencies of the order of a few Mcyc./sec., but the accuracy achieved was in general inferior to 1 % (Smith-Rose 1943), so that these measurements cannot yet be taken into account in discussions on the value of the constant.*

The experiment described in this paper presents fewer technical difficulties than any method previously described. It consists in the measurement of the frequency of electrical resonance of a hollow copper cylinder, and the calculation of the frequency from electromagnetic theory. It is similar in principle, therefore, to the experiment of Rosa & Dorsey but also has some points in common with the velocity of light experiments in that the only quantities measured are the dimensions of the apparatus and a frequency. The frequency measured is, however, that of the electromagnetic wave itself and not that of a modulation imposed on the wave. Moreover, the frequency of the oscillations is of the order of 3×10^9 cyc./sec. compared with 5×10^{14} cyc./sec. for light vibrations. The dimensions and the frequency can be alined by setting to resonance with a precision of a few parts in 10^6 so that the accuracy of observation is of the same order as that of the measurement of the dimensions and frequency. Little is gained, therefore, by making a large number of determinations, and the authors feel justified in claiming a rather higher accuracy for a single measurement than is claimed by previous workers for their averaged results. The result obtained is 299,792 km./sec. This value is 1 km./sec. lower than that which has already appeared in Essen (1947), but the difference arises only from the fact that as the work proceeded the calculations were made to more significant figures than appeared to be warranted at first.

---

* The accuracy has been greatly increased in works published since this was written and the probable error is now of the order of 1 part in 10^6 (Jones 1947; Smith, Franklin & Whiting 1947).

The experiment arose out of work on cavity resonator wave-meters, and some deductions concerning the velocity of propagation were made in a previous paper (Essen 1946). The work has been done concurrently with an already full programme, and except for the cavity resonator itself existing equipment has been used. As the experiment progressed it became clear that further work with different modes of resonance and with cavities of different types would be of interest, particularly as the final result obtained is 16 km./sec. higher than the previously accepted value of 299,776 km./sec. Such work would provide further experimental evidence concerning the extent of the different sources of possible error.

Plans have been made to continue the work as the opportunity arises, but in the meantime it is thought that the measurements already made are of sufficient interest to merit publication.

2. THEORY OF THE METHOD

The resonant frequency of a right circular hollow cylinder closed at both ends is given by Sarbacher & Edson (1943) as

$$f_{lmn} = v\sqrt{\left[\left(\frac{r}{\pi D}\right)^2 + \left(\frac{n}{2L}\right)^2\right]}, \tag{1}$$

in which f_{lmn} is the frequency in cyc./sec. of the mode designated by the suffix, v is the velocity of propagation in cm./sec. in the medium ($v = 1/\sqrt{(\mu\epsilon)}$), D is the diameter and L the length of the cylinder in cm., n is a whole number having the same value as the third suffix in the mode designation and r is the mth root of the Bessel equations $J_l(x) = 0$ for E modes and of $J'_l(x) = 0$ for H modes. If the cavity is evacuated $\mu = \mu_0$ and $\epsilon = \epsilon_0$ and the velocity is that for free space. We have then

$$v_0 = \frac{f_{lmn}}{\sqrt{\left[\left(\frac{r}{\pi D}\right)^2 + \left(\frac{n}{2L}\right)^2\right]}}. \tag{2}$$

The modes of resonance which were finally used in the experiment were the E_{010} and E_{011} and for both of these modes the value of r is 2·404825, n being 0 and 1 respectively.

The above formula is derived on the assumption that the walls of the cavity are perfectly conducting. In practice, owing to the finite conductivity of the walls, the magnetic field penetrates them to some depth, and the field configuration therefore differs from that assumed in the case of a perfect resonator. No complete solution has yet been obtained for this case, but analogy with electrical circuits resonating at low frequencies suggests that the effective inductance of the system is increased as a result of the field inside the walls and that the resonant frequency is therefore reduced. Approximate expressions have been derived for the effect on the resonant frequency. Bernier (1946), for example, has shown that if the propagation constant in the lossless guide is β ($= 2\pi/\lambda = 2\pi f/v$, where λ is the wave-length in the guide), then the propagation constant in a guide of finite conductivity is

$$\beta' = \beta\left[1 - \frac{\mu e A}{2\sqrt{2}}(1-j)\right], \tag{3}$$

where A is a function of the field and e is the skin depth which depends on the conductivity in accordance with the well-known skin-effect formulae. The imaginary part of this expression corresponds to the attenuation, but the real part corresponds to a reduction in frequency. Putting the expression in terms of frequency we have

$$f' = f\left[1 - \frac{\mu e A}{2\sqrt{2}}(1-j)\right], \tag{4}$$

and the frequency measured will be the real part of this expression, $f\left(1 - \frac{\mu e A}{2\sqrt{2}}\right)$. This can be expressed in terms of the Q of the resonator which can be measured experimentally, because

$$Q = \frac{\text{real}\,\beta'}{2\,\text{imag.}\,\beta'} = \frac{1-(\mu e A/2\sqrt{2})}{\mu e A/\sqrt{2}}, \tag{5}$$

and therefore

$$f\left(1 - \frac{\mu e A}{2\sqrt{2}}\right) = f\Big/\left(1 + \frac{1}{2Q}\right), \tag{6}$$

$(\mu e A/2\sqrt{2})$ being a very small quantity, higher powers of which have been neglected.

For v_0 we must therefore use in place of (2)

$$v_0 = \frac{f'_{lmn}(1 + 1/2Q)}{\sqrt{[(r/\pi D)^2 + (n/2L)^2]}}, \tag{7}$$

f'_{lmn} being the measured frequency and Q the measured quality factor.

The measured Q's were rather lower than the calculated values, which is a general experience with cavity resonators. The reduction is probably due to either mechanical imperfections of the surface or an increased resistivity of the skin due to the mechanical treatment it receives. In any case it seems likely that the effect causing higher resistivity must also give a greater frequency correction and that the measured value should be used.

The resonator is coupled to the source of oscillations and to a detector by small probes through holes in one of the end-walls of the cylinder, and it is important to consider the effects of this coupling on frequency. The theoretical treatment would present considerable difficulties, but previous experimental results had indicated that the degree of coupling could be reduced until the effect on frequency was negligible. It was decided, therefore, to treat this as an experimental problem and to establish that the effect was negligible within prescribed limits of accuracy.

3. DESCRIPTION OF THE APPARATUS

3·1. *General arrangement*

The apparatus is shown diagrammatically in figure 1. The cavity resonator is coupled to an oscillator and a receiver by means of the probes *A* and *B* connected to coaxial lines. The frequency of the oscillator is varied, and when it reaches a value corresponding to one of the resonant frequencies of the cavity there is a sharp increase in the amplitude of the signal, indicated on a meter in the receiver. The frequency is set to give a maximum indication and its value is then measured by means of a heterodyne wave-meter. The resonator was enclosed under a bell-jar which could be evacuated, a free air path being left in the plug and socket fittings at *A* and *B*.

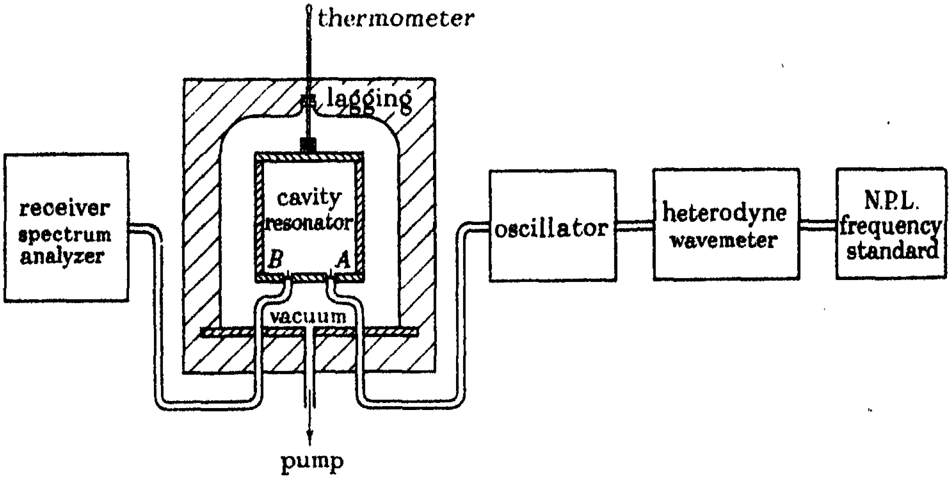


FIGURE 1. Schematic representation of the experiment.

3·2. *The cavity resonator*

A number of factors were considered in the choice of the material, shape and size of the resonator. In order to obtain a sufficiently sharp resonance the inner surface, to a depth equal to the penetration of the field, must be highly conducting, but unfortunately highly conducting metals such as copper and silver are also soft and have a high temperature coefficient of expansion. They cannot therefore be worked with the highest precision, and the measurement of the dimensions of a resonator made from such materials presents some difficulty. The alternatives of plating a hard metal or fused quartz were considered, but it was thought that the plating could not be effected with the necessary precision and that the soft metal surface would still require to be worked. Most of the advantages would therefore be annulled, and the remaining one of low-temperature coefficient was not believed to be of paramount importance as temperature was unlikely to be a limiting factor in the accuracy of measurement. It was decided, therefore, to use solid copper as the material at least for the first experiments.

The frequencies of resonance can be calculated readily by existing theoretical methods for only simple shapes of resonator, such as the right circular hollow cylinder, the sphere and the rectangular parallelepiped.

Of these the last has some advantages, since it can be built from optically flat plates and its dimensions measured by interferometer methods, but it would be more difficult to make than the cylinder, and the large number of sharp edges and joints is a disadvantage. The sphere has the advantage of freedom from sharp edges but is not easy to make with the required precision. It was decided, therefore, to use the cylindrical form for the first experiments. There is considerable latitude in the choice of the dimensions of the resonator and the resonant frequency. Oscillators are available for any frequency up to 10,000 Mcyc./sec. and the skin-effect correction decreases with increase of frequency. On the other hand, the power and frequency stability of the oscillators both decrease with increase of frequency, particularly at values much higher than 3000 Mcyc./sec. The optimum size of the cylinder from the point of view of precision construction, with existing honing equipment, was of the order of 8 cm. both in diameter and length. Such a cylinder gives low-order resonances in the region of 3000 Mcyc./sec., and it appeared therefore that this was the best frequency to use. The exact diameter and length were then chosen so that several low-order modes of resonance could be obtained in this region of frequency but spaced sufficiently to be regarded as single resonances in the theoretical treatment.

Details of the resonator are shown in figure 2. The cylinder is turned from the solid and then honed to be as uniform in dimensions as possible, the exact size not being important. The end-faces are ground square and parallel. The end-plates are ground flat and fixed to the cylinder by eight screws, an ample clearance being allowed in the holes in the plates. The coupling probes A, B pass through holes in one of the end-plates, the depth of insertion being adjustable. In the final condition no. 38 s.w.g. wire (diameter 0·0060 in. ≡ 0·0152 cm.) was used through holes of 0·03 cm. diameter. Probe A is connected to a socket in which is plugged the coaxial lead to the oscillator. If the superheterodyne receiver is used probe B is terminated in a similar socket, but if the detecting system is a crystal detector and galvanometer the crystal is connected between the probe and the socket as shown at C in the diagram. The metal piece D acts as a radio-frequency by-pass condenser and the galvanometer is plugged into the socket E. The circuit requires a direct current return lead which is provided by a wire F, the position of which can be altered to tune the crystal circuit. The insulating material G and the sockets are drilled to permit the free passage of air during the evacuation of the system.

3·3. *The receiver*

The crystal detector used in the early experiments was a standard cartridge mounted silicon crystal with a tungsten wire (type no. CV 113), and its direct current output was observed on a Tinsley type 4500 A (50 ohms) galvanometer. For the final measurements it was decided to use instead a superheterodyne receiver which was available and which proved to be slightly more sensitive. This is a special form of instrument (known as a spectrum analyzer or spectrometer) developed for observing the spectral distribution of energy in the output of transmitting valves. Its useful feature in the present application is that its sensitivity remains almost constant, without adjustment of any controls, over a small band of frequencies. The tuning controls of the receiver are set so that the resonant frequency of the cavity is near the middle of this band. Then as the oscillator frequency is varied through the resonant value a peak deflexion is observed without any retuning of the receiver.

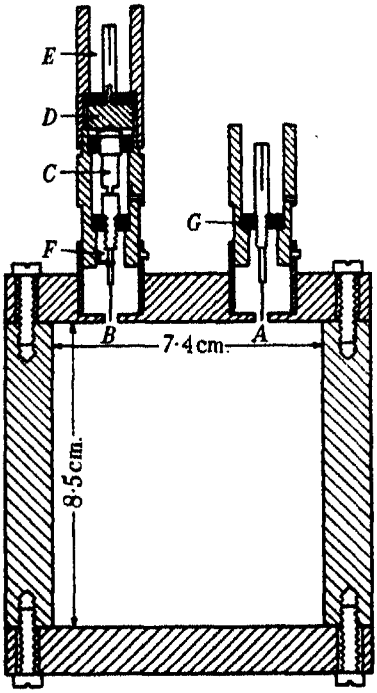


FIGURE 2. Details of the resonator.

3·4. *The heterodyne wave-meter*

This instrument and its method of use have already been described by the authors (1945). With careful operation it enables frequencies in the region concerned to be measured with an accuracy of ± 2 parts in 10^8.

3·5. *The oscillator*

The oscillator valves were chosen from those used as local oscillators in receivers, the type CV 35 being suitable for the E_{010} mode and the CV 234 for the E_{011} mode. The former was operated from a stabilized mains supply unit and the latter from batteries. The tuning mechanisms supplied with the valves were found to give a sufficiently smooth frequency variation, although it was sometimes necessary to choose a good portion of the range of the tuning mechanism by the appropriate adjustment of subsidiary tuners.

4. MEASUREMENT OF RESONANT FREQUENCIES AND QUALITY FACTORS

The assembly of resonator and bell-jar was well lagged to ensure that any temperature changes occurring were very slow, and that the value recorded on a thermometer in good thermal contact with the resonator represented its actual temperature within 0·1° C. The system was evacuated at least 12 hr. before the electrical measurements were made. The electrical equipment was also switched on some hours before it was required for use in order that the various oscillators would reach a steady temperature and the measurements could be made under the most favourable conditions.

As explained in § 3·1 the measurement consisted in the adjustment of the oscillator to the resonant frequency of the cavity and then in the measurement of this frequency in terms of the standard. It was carried out by two observers to eliminate any appreciable time interval between the two operations. Precautions were taken to ensure that the maximum deflexion in the receiver did in fact correspond to the resonant condition of the cavity.

It was established that there was no appreciable change in the output of the oscillator as the tuning was varied in the region of resonance. There are in the circuit several short lengths of cable and connecting plugs and sockets. The impedance of these and therefore the power available at the probes changes with frequency, but owing to the resistance of the cable such tuning effects were of a different order of sharpness from the cavity resonance. It was easy to arrange that for the small frequency change involved in passing through resonance there was no appreciable amplitude change due to these effects. An overall check was obtained, in preliminary experiments, by increasing the insertion of the probes in order to give a convenient deflexion on the oscilloscope of the receiver by direct coupling between them when the cavity was not in resonance. The frequency was varied a few megacycles on both sides of the resonant value, and it was verified that the amplitude of the oscilloscope deflexion remains constant as it moves along the time base on both sides of the cavity resonance.

The voltage supplies to the oscillator were set so that frequency variations due to any residual fluctuations in them were reduced to a minimum. Under good conditions the frequency band over which an audible beat note could be obtained in the heterodyne wave-meter was less than 1 part in 10^5 of the frequency, and the setting to the mid-point of the band was made with a precision of ± 2 parts in 10^6.

The Q of the resonator was measured by observing the change in frequency from the resonant value required to reduce the detected current to $1/\sqrt{2}$ of its peak value. If this change is δf then the Q is $f/2\delta f$. The value was checked by measurements made with the crystal detector and galvanometer under conditions for which the detector was known to give a square law within close limits.

5. MEASUREMENT OF DIMENSIONS

The dimensions of the resonator were measured in the Metrology Division of the National Physical Laboratory.

5·1. *Internal diameter*

The measurements were made in a temperature-controlled room with a horizontal comparator arranged for internal measurements through steel contact tips of rounded form (about 0·4 cm. radius) operating under a measuring force equivalent to 0·34 kg. weight. A suitably wrung combination of slip-gauges and end-pieces served as the reference basis for internal comparison. Differences between this and any particular diameter of the tube were indicated on the optical scale of the comparator which could be read directly to $1\cdot3\mu$ and by estimation to $0\cdot3\mu$ (0·00001 in.). The measurements were made over eight symmetrically disposed diameters at seven positions along the axis of the cylinder.

The accuracy of determination of the mean measured diameter was estimated at $\pm 1\cdot3\,\mu$ for the preliminary measurements which were carried out as normal routine tests. To improve this for the final measurement the following modifications were made in the technique of measurement:

(*a*) The comparator was screened by means of a glass panel so that temperature effects due to the presence of the observer were minimized as much as possible.

(*b*) An improved form of reference standard was used. This was of the 'box' type and consisted of a firmly wrung combination of two piles of high-quality slip-gauges not differing in length by more than about $0\cdot03\mu$, between two lapped end-pieces, thus constituting a rigid 'box', or frame for which the reference dimension was well established. Each slip-gauge employed in the combination had previously been calibrated by interferometer measurements in terms of wave-lengths of light to an accuracy equivalent to $0\cdot03\mu$.

(*c*) Two diameters at the centre of the copper cylinder and in directions at right angles to one another were accurately measured in terms of the reference standard.

(*d*) The temperatures of the copper cylinder and the hardened steel standard were observed and the results of the intercomparison were reduced to 20° C, using thermal coefficients of expansion of 12×10^{-6} and 17×10^{-6} for hardened steel and copper respectively.

(*e*) Having established two basic diameters all other diameters of the cylinder were compared with first one and then the other of the basic diameters. The effect of any temperature variations on the results obtained during each of these long series of comparisons was thereby eliminated provided any changes in temperature were reasonably the same for all parts of the cylinder. The mean measured diameter derived from each of the two series agreed to $0\cdot5\mu$.

(*f*) Theoretical and experimental investigations were made of the different effects of the compressive force exerted by the steel measuring tips of the comparator on hardened steel and copper surfaces. The results obtained by calculation and experiment agreed to $0\cdot03\mu$ and showed that the difference could be compensated by reducing the observed value of internal diameter of the cylinder as derived from the standard by $0\cdot27\mu$.

5·2. *Length*

Length measurements were made by means of a vertical comparator having a rounded contact measuring tip and an optical scale similar to that of the horizontal comparator. One end of the cylinder was placed in contact with the horizontal platen of the comparator under the contact tip, and the height of the other end of the cylinder was measured at various positions by comparison with a wrung combination of reference standard slip-gauges. Under these conditions of measurement it was considered that the compressive effect of the single contact tip was negligible in relation to the general accuracy of the length determination.

The overall length of the tube was thus measured at eight positions symmetrically spaced around the cylinder, and at each of these positions three measurements were made, one near the bore, one at the mid-point of the wall thickness and one near the outside of the tube. The measurements are considered to be reliable to $\pm 0\cdot 8\mu$.

Some tests were made to show whether the diameter and length were affected by screwing on the end-plates, but within the accuracy of these measurements ($\pm 0\cdot 8\mu$) there was no evidence of deformation.

6. Experimental results

6·1. *Preliminary measurements*

The method of coupling used favours the excitation of the E_{01n} modes but the H_{011} and H_{111} modes can be excited if the probes are bent round to form small loops orientated so as to give a linkage with the field associated with the particular mode. To gain experience of the most suitable experimental conditions preliminary measurements were therefore made of the resonant frequencies of a number of different modes.

From the frequencies the values for velocity were calculated and the results are included in table 1, but they are not considered to be more accurate than 6 parts in 10^6 for the following reasons:

(*a*) The metrological examination of the cylinder revealed larger variations of diameter than anticipated, and the measurements themselves were not made with the maximum possible accuracy, the uncertainties being given at the head of table 1. The end-faces were not ground flat and the length measurement is therefore averaged to allow for this, the deviations being of the order of $\pm 5\mu$.

(*b*) The conditions for the frequency measurement were not ideal. For the E_{012}, H_{011} and H_{111} modes the coupling probes were inserted about 1 or 2 mm., and in the case of the two latter modes were formed into small loops. The size of the coupling holes was 0·15 cm.

Table 1. (Preliminary measurements.) Resonant frequencies of the copper cylinder and calculated values of velocity

Mean diameter of resonator $7\cdot 39759 \pm 0\cdot 00012$ cm. Variation $\pm 0\cdot 00045$ cm.
Length of resonator $8\cdot 55844 \pm 0\cdot 0003$ cm.

mode of resonance	correction factor $(1+1/2Q)$	constant r	frequency f_0' (Mcyc./sec.)	v_0 (km./sec.)
E_{010}	1·000028	2·404825	3102·12	299,797
E_{011}	1·000035	2·404825	3562·38	299,798
E_{012}	1·000030	2·404825	4678·81	299,785
H_{011}	1·000015	3·831706	5243·96	299,799
H_{111}	1·000031	1·841184	2950·78	299,777

(*c*) As a result of these measurements it was decided to concentrate on the E_{010} and E_{011} modes, which could be detected readily with the probes flush with the wall of the cavity. The size of coupling hole was reduced. It was at the same time decided to regrind the cylinder to improve its uniformity of diameter and length.

6·2. *Effect of size of coupling hole and the penetration of the probes*

It was necessary to establish experimentally the effect on frequency of the presence of the coupling holes and probes. First measurements were made with three sizes of coupling hole, the probes being in each case as nearly flush as possible to give an adequate indication. The results are given in table 2. It was concluded from these results that any residual effect on frequency was less than 0·01 Mcyc./sec., i.e. less than 3 parts in 10^6.

TABLE 2

mode	size of coupling hole (cm.)	insertion of probe (cm.) (estimated by eye)	resonant frequency (Mcyc./sec.)
E_{010}	0·03	<0·01	3101·246
	0·07	flush	3101·251
	0·16	flush	3101·242
E_{011}	0·03	<0·01	3563·798
	0·07	flush	3563·787
	0·16	flush	3563·71

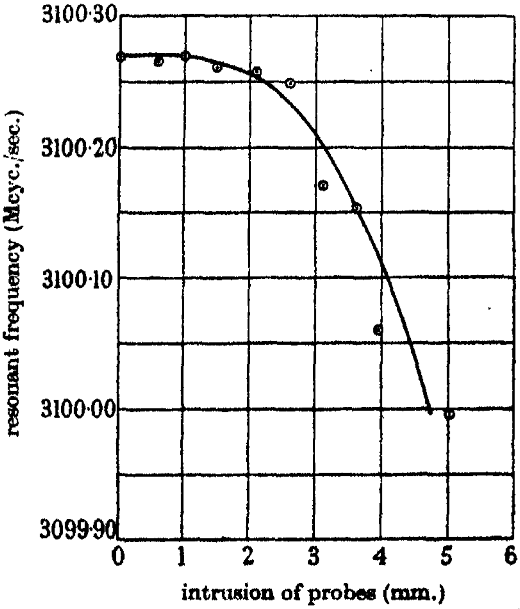


FIGURE 3. E_{010} mode (in air).

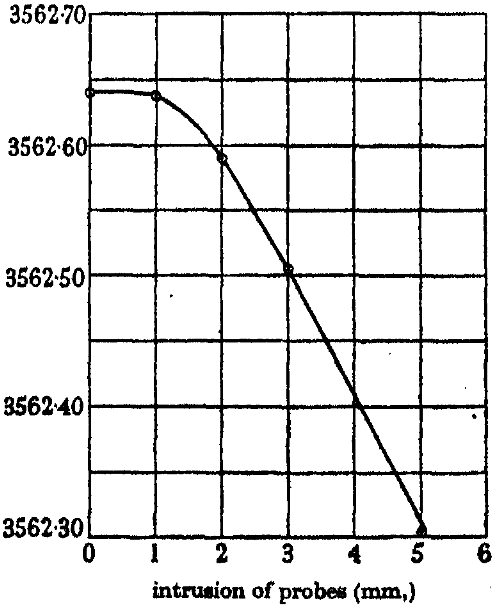


FIGURE 4. E_{011} mode (in air).

FIGURES 3 and 4. Effects of probe intrusion. Diam. of holes 0·03 cm. Diam. of probes 0·015 cm.

The effect of probe insertion is shown in figures 3 and 4, and from these results it is concluded that the residual effect of the probes is again less than 0·01 Mcyc./sec.

The Q of the resonator for the two modes had already reached its maximum value with the coupling considerably closer than that finally used. The values are 18,000 and 14,000 respectively for the E_{010} and E_{011} modes as compared with the theoretical values of 21,000 and 17,600. The ratios of experimental to theoretical values are as great as those usually obtained for cavity resonators.

6·3. *Final measurements*

The final results for the velocity are based on a series of metrological measurements made between two series of electrical measurements. The detailed dimensions are given in tables 3 and 4. The cylinder is seen to be slightly oval in shape, the difference between the maximum and minimum diameters being 2μ (3 parts in 10^5). Since the field for the modes employed is symmetrical about the axis, the effective diameter was taken as the mean value. The length at the inner edge of the cylinder was slightly greater than that at the outer edge and varied smoothly round the circumference by a total of $2{\cdot}2\mu$. The effective length was therefore taken as the mean value at the inner edge.

TABLE 3. DIAMETER OF THE RESONATOR EXPRESSED AS THE DIFFERENCE FROM THE MEAN VALUE OF 7·39957 CM. UNIT 0·00001 CM. ($0{\cdot}1\mu$)

axial position of measurement from one end (cm.)	measured diameter at 20° C in different diametrical planes							
	1	2	3	4	5	6	7	8
0·13	−15	− 6	+ 1	+ 1	+ 7	+ 3	0	−10
1·3	−20	−11	− 4	− 2	+ 1	+ 2	− 4	−16
2·8	−20	−13	−17	− 2	+ 9	+ 7	− 4	−20
4·3	−15	− 9	− 6	+ 3	+ 8	+10	− 3	−14
5·8	− 2	+ 6	+13	+13	+27	+24	+13	0
7·3	− 3	− 3	+ 6	+10	+14	+ 9	+ 7	0
8·4	+ 7	− 1	+ 5	+ 8	+10	+ 8	+ 2	+ 9

TABLE 4. LENGTH OF THE RESONATOR EXPRESSED AS THE DIFFERENCE FROM 8·53637 CM. (MEAN OF INNER EDGE VALUES). UNIT 0·00001 CM. ($0{\cdot}1\mu$)

position of measurement	mean overall length at 20° C		
	near inner edge	centre	near outer edge
1	− 8	−13	−16
2	+ 5	+ 2	− 8
3	+ 9	+ 6	+ 1
4	+ 9	+ 9	+ 3
5	+ 7	+ 2	− 3
6	− 2	− 5	−12
7	−13	−14	−17
8	− 9	−13	−17

The measured values of resonant frequency and the values of velocity calculated from the resonant frequencies and dimensions are given in table 5. The frequencies were measured using the minimum coupling conditions as follows:

Size of coupling holes	0·03 cm.
Diameter of probes	0·015 cm.
Insertion of probes	< 0·01 cm.

Resonance was detected on the spectrum analyzer as described in § 3·3.

TABLE 5. FINAL VALUES FOR RESONANT FREQUENCIES OF THE E_{010} AND E_{011} MODES AND THE DEDUCED VALUES OF THE VELOCITY OF PROPAGATION

Length of resonator 8·53637 ± 0·00008 cm. Diameter of resonator 7·39957 ± 0·00003 cm.

date	mode of resonance	correction factor $(1+1/2Q)$	constant r	measured frequency f'_{lmn} (Mcyc./sec.)	v_0 (km./sec.)
2. x. 46	E_{010}	1·000028	2·404825	3101·25	299,793
2. x. 46	E_{011}	1·000035	2·404825	3563·80	299,791
25. x. 46	E_{010}	1·000028	2·404825	3101·28	299,796
25. x. 46	E_{011}	1·000035	2·404825	3563·77	299,789

mean value of velocity 299,792 km./sec.

The estimated maximum errors due to various causes are listed below:

(1) Setting of the frequency to resonance and measurement of the frequency	$0{\cdot}4\times10^{-5}$
(2) Uncertainty of temperature of the resonator	$0{\cdot}2\times10^{-5}$
(3) Dimensional measurements	$0{\cdot}3\times10^{-5}$
(4) Residual effects of coupling holes and probes	$0{\cdot}6\times10^{-5}$
(5) Non-uniformity of the resonator	$1{\cdot}0\times10^{-5}$
(6) Uncertainty of Q	$0{\cdot}5\times10^{-5}$
Estimated maximum error	3×10^{-5}

The uncertainties due to the first three causes can be established from the accuracy of repetition and a knowledge of the band-width of the receiver and frequency stability of the oscillator. The errors due to the other causes cannot be estimated with such certainty. That due to the coupling is estimated from the results given in § 5·2. As regards the non-uniformity of the resonator it will be seen from tables 3 and 4 that the maximum deviations of the dimensions from the mean are $+2{\cdot}7$ and $-2{\cdot}0\mu$, and that the variations occur in a smooth manner. It is estimated therefore that the average dimensions are known to $0{\cdot}5\mu$. Previous results (Essen 1946) obtained with much less uniform resonators support the assumption that an average value can be taken, although it cannot be assumed that the non-uniformity of dimensions affects different modes of resonance in the same way. An overall error of 1×10^{-5} has therefore been allowed for these uncertainties.

The total value of the correction due to the finite resistance of the walls of the resonator is approximately 3×10^{-5}. An uncertainty is introduced because of the imperfections in the surface of the resonator and a lack of knowledge of the precise path of the current. It is thought that these effects are to some extent taken account of by using the measured value of Q in the calculations. The value of Q is accurate to about 5 %, and the overall error has been estimated at $0{\cdot}5 \times 10^{-5}$.

The results obtained before and after the metrological measurements do not give quite such good agreement as was usually obtained during the preliminary measurements on dismantling and reassembling the resonator. However, in view of the importance of associating the frequency measurements closely in time with the measurement of dimensions, the values were accepted and their mean taken as giving the final result. The difference between the values of velocity obtained for the two modes is also rather larger than expected. The uncertainty in the dimensions could contribute to this, since the length and diameter affect the resonant frequencies to different extents, and the residual errors due to coupling and non-uniformity could also have different effects on the two modes.

The work described above was started in the Radio Division of the National Physical Laboratory as part of the programme of the Radio Research Board and completed in the Electricity Division as part of the research programme of the National Physical Laboratory. This paper is published by permission of the Department of Scientific and Industrial Research.

The authors wish to acknowledge the assistance of Mr H. Barrell of the Metrology Division who was responsible for the measurement of the dimensions of the resonator, and of Mr J. W. Simmons of that Division who was responsible for its construction.

REFERENCES

Anderson, W. C. 1941 *J. Opt. Soc. Amer.* **31**, 187.
Bernier, J. 1946 *Onde Elect.* **26**, 305.
Birge, R. T. 1941 *Physics*, **8**, 90.
Essen, L. 1947 *Nature*, **159**, 611.
Essen, L. 1946 *J. Instn Elect. Engrs*, **93**, part IIIA, no. 9, p. 1413.
Essen, L. & Gordon-Smith, A. C. 1945 *J. Instn Elect. Engrs*, **92**, 291.
Jones, F. E. 1947 *J. Instn Elect. Engrs*, **94**, part III, p. 399.
Michelson, A. A., Pease, F. G. & Pearson, F. 1935 *Astrophys. J.* **82**, 26.
Rosa, E. B. & Dorsey, N. E. 1907 *Bull. U.S. Bur. Stand.* **3**, 433.
Sarbacher, R. I. & Edson, W. A. 1943 *Hyper- and ultra-high frequency engineering.* London: Chapman and Hall.
Smith-Rose, R. L. 1943 *J. Instn Elect. Engrs*, **90**, part I, no. 25, p. 29.
Smith, R. H., Franklin, E. & Whiting, F. B. 1947 *J. Instn Elect. Engrs*, **94**, part III, p. 391.

The constitution of the aluminium-rich aluminium-cobalt-iron alloys, with reference to the role of transitional elements in alloy formation

By G. V. Raynor and M. B. Waldron

Department of Metallurgy, The University of Birmingham

(*Communicated by W. Hume-Rothery, F.R.S.—Received* 3 *January* 1948)

[Plate 9]

The behaviour of transitional metals in aluminium-rich alloys is of theoretical interest. From a knowledge of the system aluminium-iron-nickel (Raynor & Pfeil 1946–7*a*) it has proved possible to predict the general form of the equilibrium diagram for the aluminium-rich aluminium-cobalt-nickel alloys (Raynor & Pfeil 1946–7*b*). Similar predictions may be made for the aluminium-iron-cobalt alloys. The present paper describes the results of an examination of these alloys undertaken to test the predictions. Using micrographic and X-ray methods, isothermal sections have been established at 640, 600 and 550° C. Only the two intermetallic phases Co_2Al_9 and $FeAl_3$ have been recognized. Vertical sections across the aluminium-rich corner of the ternary model have been established by thermal analysis, and the fields of primary separation have been determined. Analysis of extracted primary crystals has shown that Co_2Al_9 dissolves appreciable amounts of iron, while $FeAl_3$ dissolves a similar amount of cobalt. In each case the solution proceeds by atomic replacement. The results are discussed in their relation to the theoretical considerations involved, and it is shown that close analogies exist between the alloys of aluminium with iron and nickel, with iron and cobalt, and with cobalt and nickel.

1. Introduction

The present work forms part of a research programme on aluminium-rich alloys containing iron, cobalt and nickel, undertaken in order to study the role of transitional metal solutes in this class of alloy. The constitutions of the component binary systems, and of the ternary system aluminium-iron-nickel, are well established, and Bradley & Taylor (1940) have demonstrated the isomorphism of the compound Co_2Al_9 with $FeNiAl_9$. They also suggested that there may be a plane in the model of the system aluminium-iron-nickel in which conditions are analogous to those in the binary aluminium-cobalt system, which implies that two atoms of cobalt are approximately equivalent to one atom of iron plus one atom of nickel in the ternary alloys.

In recent papers (Raynor & Pfeil 1946–7*a*, *b*; Raynor 1944; Raynor & Little 1945 and Raynor 1945), contributions have been made towards a theory of the role of transitional metal solutes in aluminium-rich alloys. According to this theory, transitional metal atoms, when alloyed with aluminium, absorb from the structure as a whole a number of electrons per atom corresponding to the total available vacancies in their atomic orbitals. If the theories of Pauling (1938) are accepted, then the available vacancies for chromium, manganese, iron, cobalt and nickel are respectively 4·66, 3·66, 2·66, 1·71 and 0·61 per atom. The application of this principle enables several analogies to be established between different aluminium-rich binary and ternary alloy systems containing transitional metals. In particular the isomorphism of Co_2Al_9 and $FeNiAl_9$ may be attributed largely to the close similarity between the number of vacancies in the atomic orbitals for two atoms of cobalt and for one atom of iron plus one atom of nickel.

Raynor & Pfeil (1946–7*a*) re-examined the equilibrium relations and other features in the aluminium-iron-nickel system, and further suggested that it was possible to regard the constitutional model as built up from the aluminium-cobalt-iron and aluminium-cobalt-nickel diagrams placed side by side. This suggestion was partially confirmed by Raynor & Pfeil (1946–7*b*), who showed that there is a very close analogy between the aluminium-cobalt-nickel system, and the aluminium-iron-nickel alloys which contain nickel in excess of the equiatomic ratio Fe : Ni = 1 : 1. This paper described the results of a corresponding examination of the system aluminium-iron-cobalt. Predictions of the general constitutional features expected are summarized below:

(*a*) In the aluminium-iron-nickel alloys in which the ratio of iron to nickel atoms is greater than 1 : 1, equilibrium in the aluminium-rich corner involves only the primary solid solution, $FeAl_3$* and $FeNiAl_9$. No ternary compound would therefore be expected to enter into equilibrium with the primary solid solution in the aluminium-iron-cobalt alloys, but only the phases $FeAl_3$ and Co_2Al_9.

(*b*) In the compound $FeNiAl_9$, nickel is replaced by iron, atom for atom, to a limiting composition of 17·13 weight % iron and 14·88 weight % nickel, which is a little nearer to the aluminium-iron axis of the ternary diagram than the theoretical composition $FeNiAl_9$. According to our theory of the role of transitional metals when alloyed with aluminium the electron : atom ratios corresponding to this limiting composition and to Co_2Al_9 are almost the same. It would therefore be expected that the solubility of iron in Co_2Al_9 would be limited.

(*c*) Since the solubility of nickel in $FeAl_3$ is appreciable, it would be expected that an appreciable solubility of cobalt in $FeAl_3$ would also be observed, the replacement occurring in such a way that one atom of cobalt replaced one atom of iron.

2. THE BINARY SYSTEMS

The aluminium-rich portions of the equilibrium diagrams of the systems aluminium-cobalt and aluminium-iron have been previously described (Raynor & Pfeil 1946–7*a*, *b*), and are reproduced in figures 1 and 2 of this paper. The important features are:

(*a*) Aluminium-cobalt. The maximum solubility of cobalt in aluminium is less than 0·02 %. At 1 % cobalt and 657° C, the aluminium-rich solid solution and the compound usually denoted Co_2Al_9 separate from the melt as a eutectic mixture.

(*b*) Aluminium-iron. The maximum solubility of iron in aluminium is less than 0·02 %, while at 1·7 % iron and approximately 655° C the compound $FeAl_3$ and the aluminium-rich solid solution separate as a eutectic mixture. At some temperature below 600° C, however, $FeAl_3$ decomposes to form the compounds Fe_2Al_7 and Fe_2Al_5 (Bradley & Taylor 1938), from which it is virtually indistinguishable micrographically.

---

* Bradley and his colleagues have shown that $FeAl_3$ decomposes into Fe_2Al_7 and Fe_2Al_5 on slow cooling to room temperature. For consistency in nomenclature, however, the term $FeAl_3$ is in general retained in this paper.

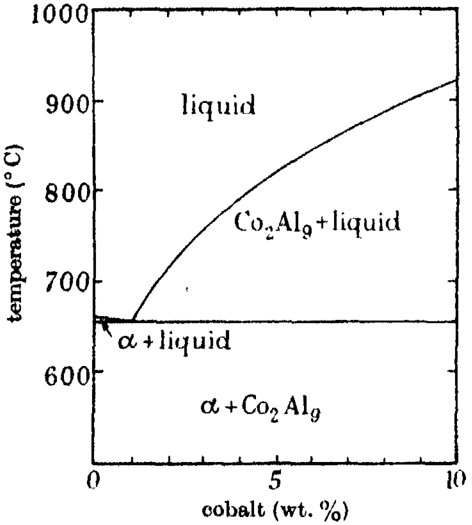


FIGURE 1. Equilibrium diagram for aluminium-cobalt system.

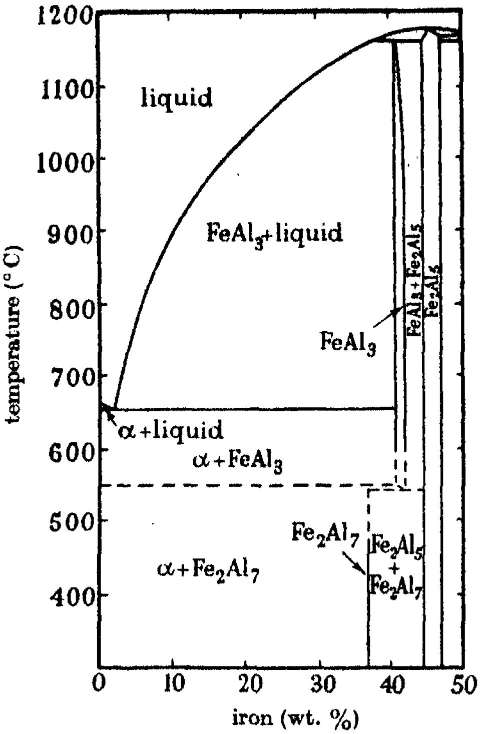


FIGURE 2. Equilibrium diagram for aluminium-iron system.

3. GENERAL EXPERIMENTAL PROCEDURE

(a) *Materials used*

Alloys for all experiments were prepared from super-pure aluminium and the binary master alloys listed in table 1. The alloy of highest cobalt content was kindly presented by the Mond Nickel Company, Ltd., while for the remaining alloys grateful acknowledgement is due to the British Aluminium Company Ltd. Alloy 1 in table 1 was reserved for the preparation of specimens for annealing work in view of its superior quality.

TABLE 1

element (wt. %)	aluminium-cobalt alloys 1	2	3	aluminium-iron alloy 4
Co	2·50	4·65	11·0	—
Fe	0·005	0·020	0·05	5·72
Si	0·0035	0·020	0·05	0·0035
Cu	0·001	0·025	—	0·0035
Al	remainder	remainder	remainder	remainder

(*b*) *Experimental methods*

Alloys were prepared in alumina-lined crucibles, cast into chill moulds, and annealed in evacuated hard-glass tubes. Reaction between specimens and the glass was prevented by enclosing the former in alumina sheaths. Annealing treatments were carried out at temperatures controlled automatically to within ± 1·5° C. of those desired. On the completion of annealing treatments, all alloys were rapidly quenched in water.

For thermal analysis experiments, alloys were cooled at an automatically controlled rate of 1·5° C per min., in a furnace from which incidental temperature fluctuations were eliminated by the use of a heavy copper lining. Continuous stirring was employed throughout to minimize super-cooling. Arrest temperatures were measured with a calibrated platinum/platinum-13 %-rhodium thermocouple, used in conjunction with a deflexion potentiometer.

Alloys intended for the extraction of primary crystals of the intermetallic compounds were cooled at a similar rate to that employed in the thermal analysis, but were not stirred. The solid ingots were examined microscopically, and primary crystals were extracted by anodic solution of the matrix in hydrochloric acid solution. Samples of clean, good quality crystals were picked out by hand from the residue, and submitted to Messrs Johnson Matthey and Company, Ltd, for analysis.

Microscopical examination of annealed alloys was carried out by standard techniques; the clearest distinction between Co_2Al_9 and $FeAl_3$ was given by immersion for 45 sec. in 10 % caustic soda solution, used cold, after standing for at least 24 hr. before use. The microstructure of an alloy containing Co_2Al_9 (dark), $FeAl_3$ (light) and chilled liquid is shown in figure 8, plate 9 (see §5).

Standard techniques were also employed for the X-ray examination of certain samples of extracted primary crystals.

In the text of this paper, individual alloys are denoted by reference to the amounts of cobalt and iron present, expressed as percentages by weight. Thus, alloy 0·50/2·00 contained 0·50 % cobalt and 2·00 % iron.

4. CONSTITUTION IN THE SOLID STATE

The constitution of the system up to a cobalt content of 11 weight % and an iron content of 5·7 weight % was determined using 29 alloys. Specimens were annealed at 640, 600 and 550° C. They were examined at suitable stages in the course of their heat-treatment, and annealing was continued until no further change in microstructure occurred. The disposition of the phase fields obtained thus represents true equilibrium conditions. Critical micro-sections of alloys used to establish the 600° C isothermal were analyzed; the nominal compositions of the remaining alloys were used in fixing the phase boundaries at 640° C.

(*a*) The 640° C isothermal. The structures of the alloys annealed for 14 days at 640° C are summarized in figure 3. No evidence of any ternary compound was obtained.

(*b*) The 600° C isothermal. The structures of the alloys annealed at 600° C are given in figure 4. All alloys which determine phase-boundaries received a minimum annealing time of 35 days, although no structural change was detected on re-annealing any alloy which had received 17 days treatment. The two-phase ($\alpha + FeAl_3$) and ($\alpha + Co_2Al_9$) fields are narrower than at 640° C, while the three phase-triangle has widened.

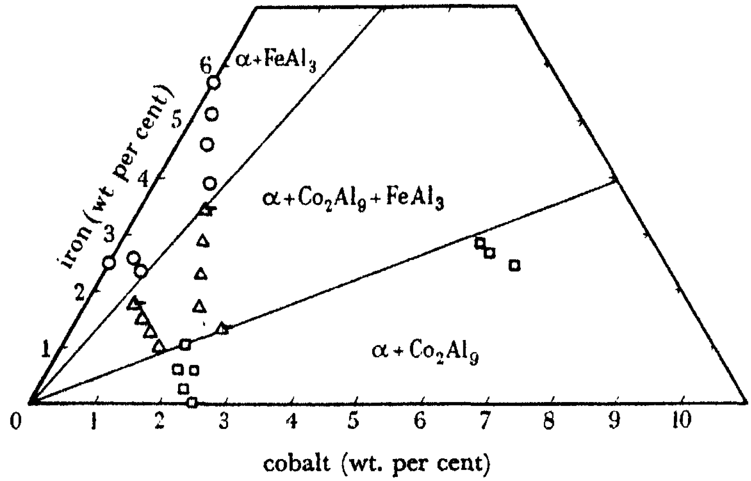


Figure 3. Constitution at 640° C.

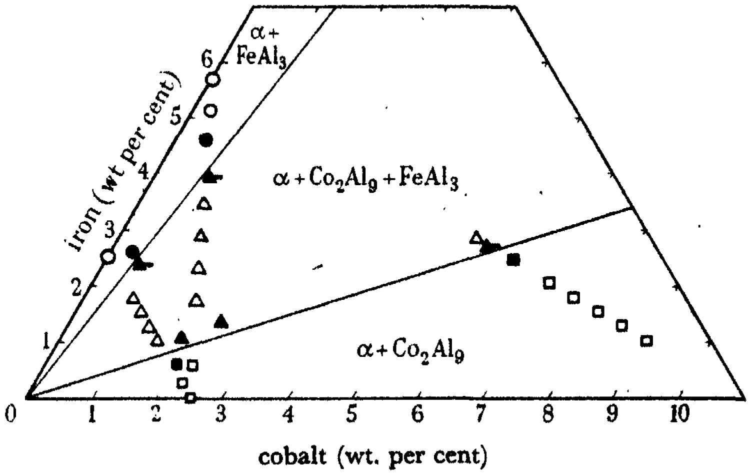


Figure 4. Constitution at 600° C.

Key to figures 3 and 4. ○, $\alpha + FeAl_3$. △, $\alpha + Co_2Al_9 + FeAl_3$. △, alloys with trace of third phase. □, $\alpha + Co_2Al_9$. ●▲▲■, analyzed alloys.

(c) The 550° C isothermal. Alloys annealed for a period of 8 weeks at 550° C showed no change in microstructure after the fifth week. The isothermal diagram for this temperature differs from that at 600° C only by a slight narrowing of the ($\alpha + Co_2Al_9$) field. A diagram of the constitution at 550° C is not included, in view of the very small differences from the 600° C isothermal (figure 4).

5. THERMAL ANALYSIS

Figures 5, 6 and 7 show vertical sections through the ternary model at 98·75, 98·5 and 97·5 % of aluminium respectively.

The binary aluminium-cobalt eutectic temperature was determined as 657·1° C, in good agreement with the value of 657° C reported by Fink & Freche (1932). For the aluminium-iron eutectic temperature, arrests were observed at 654·5 and 655·1° C, again in substantial agreement with previous work (Archer & Fink 1928; Dix 1925).

Well-developed ternary eutectic arrests were obtained with nine alloys. The most probable value for the ternary eutectic temperature is 653·9 ± 0·3°C.

The other important features of the cooling curve work may be summarized as follows:

(*a*) The vertical section at 98·75 % aluminium (figure 5). The position of the eutectic valley between the solid solution and Co_2Al_9 lies at 0·72 % cobalt and 0·53 % iron, within ± 0·06 % of either element. Alloy 0·625/0·625, examined micrographically, showed primary crystals of the aluminium-rich solid solution. A similar examination of alloy 0·75/0·50 revealed no evidence of primary separation of either constituent; in alloy 0·875/0·375, however, small crystals of primary Co_2Al_9 were detected.

(*b*) The vertical section at 98·5 % aluminium (figure 6). The eutectic valley between the primary solid solution and Co_2Al_9 liquidus surfaces now occurs at 0·55 % cobalt and 0·95 % iron.

(*c*) The vertical section at 97·5 % aluminium (figure 7). This section extends from the field in which Co_2Al_9 is primary to that of primary $FeAl_3$, without crossing any region in which the aluminium-rich solid solution separates as primary crystals. Location of the point *B* on thermal evidence only was rendered uncertain owing to the weak character of the arrests in this region. Further information was therefore obtained micrographically at temperatures between the liquidus and the eutectic horizontal. A typical area containing Co_2Al_9 and $FeAl_3$ in a matrix of chilled liquid, as seen etched, is shown in figure 8, plate 9. The microstructures of six alloys, after annealing for four hours, are included in figure 7, and support the interpretation of the thermal arrests. The boundary between the two primary fields occurs at 0·75 % cobalt and 1·75 % iron.

(*d*) The surfaces of primary separation. Interpolation of the cooling curve data permits an estimate to be made of the composition of the ternary eutectic, and of the disposition of the surfaces of primary separation. The eutectic valley between the primary solid solution and primary Co_2Al_9 is located by the composition of the binary eutectic itself, and by the intersections of the liquidus curves in figures 5 and 6. The boundary between the primary $FeAl_3$ and Co_2Al_9 fields is fixed at one point by the intersection of the liquidus curves in figure 7, and at a point more remote from the aluminium-rich corner of the diagram by evidence obtained from slowly-cooled alloys (see § 6). Over such a small composition range, it is unlikely that there is any marked curvature of this boundary, which may therefore be represented approximately by a straight line through these two points. This line intersects the

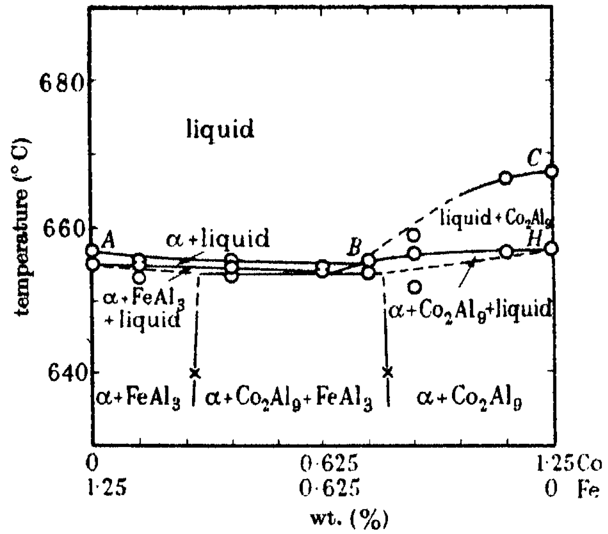


FIGURE 5. Vertical section through ternary model at 98·75 % aluminium.

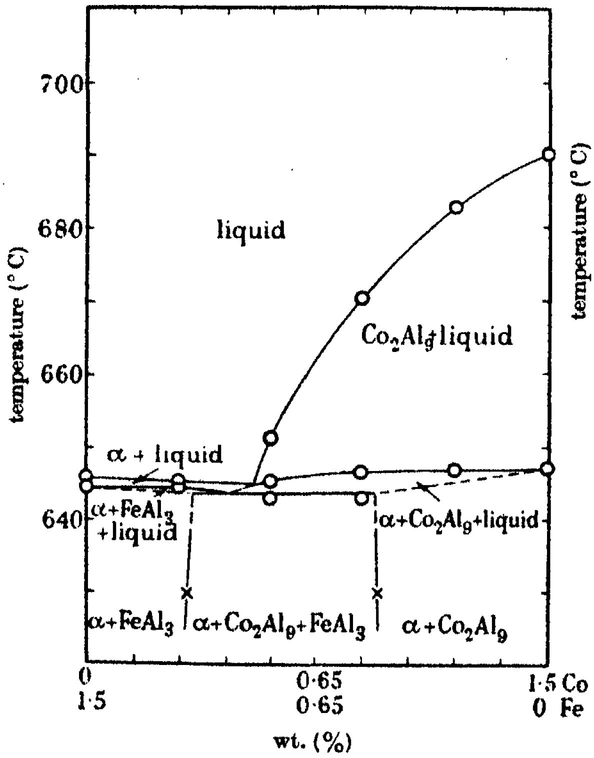


FIGURE 6. Vertical section through ternary model at 98·5 % aluminium.

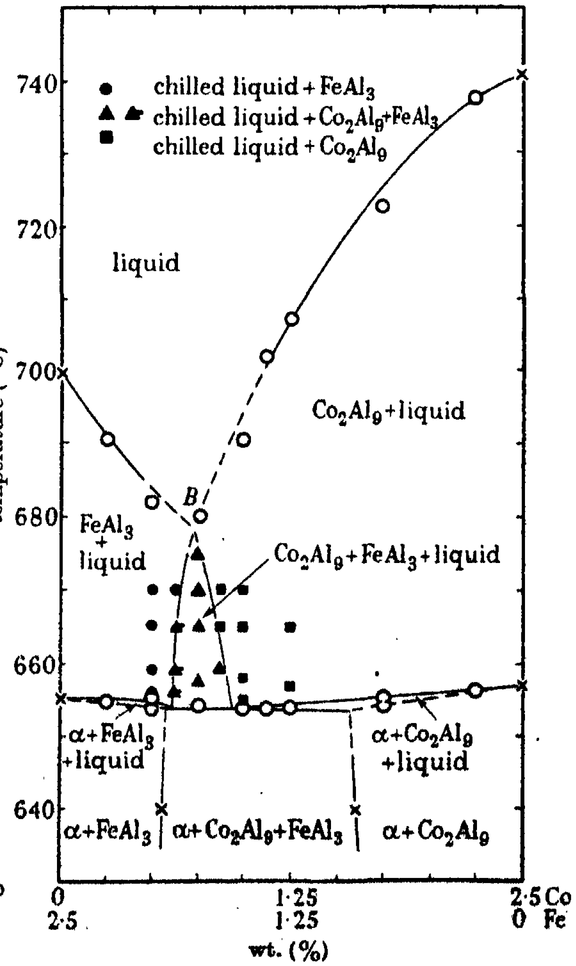


FIGURE 7. Vertical section through ternary model at 97·5 % aluminium.

FIGURE 8. (× 250.)

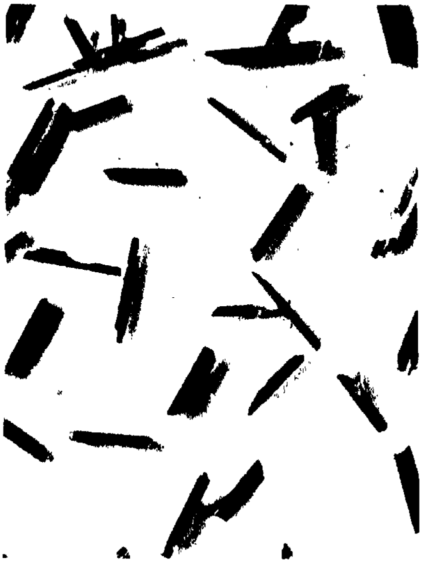


FIGURE 10. (× 7.)

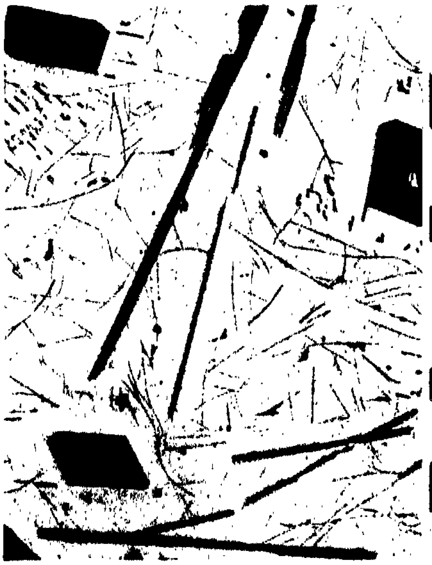


FIGURE 11. (× 75.)

FIGURE 12. (× 7.)

α/Co_2Al_9 eutectic valley at 0·35 % cobalt and 1·55 % iron. These figures give the most probable composition of the ternary eutectic point, within extreme estimated limits of ± 0·1 % cobalt and ± 0·2 % iron. Figure 9 shows the projection of the fields of primary separation on to the base of the ternary model; the broken lines represent isothermal contours obtained by graphical interpolation of the cooling curve results.

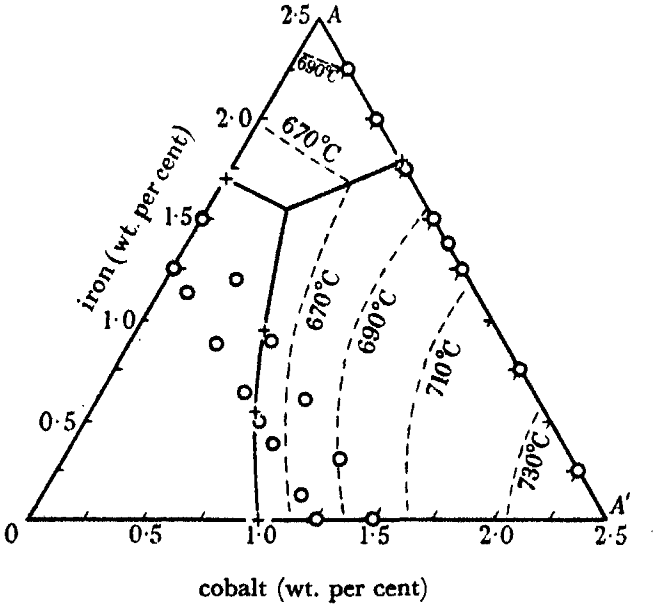


FIGURE 9. Surfaces of primary separation.

6. ANALYTICAL RESULTS FOR THE EXTRACTED RESIDUES

Six alloys, with compositions lying on a line stretching from 2·5 % cobalt to 5·7 % iron, were slowly cooled, and samples of primary crystals were extracted as described. The crystals from alloys containing 0·57, 1·14 and 1·71 % iron were prismatic in habit, as shown in figure 10, plate 9. Microscopical examination of the parent ingots showed that these crystals possessed the etching characteristics of Co_2Al_9. The microstructure of alloy 1·50/2·29 showed that primary crystals of $FeAl_3$ were present, in addition to those of primary Co_2Al_9; an area of this alloy is shown in figure 11, plate 9. The occurrence of two kinds of primary crystals in the same slowly cooled alloy has been observed previously (Raynor & Pfeil 1946–7*a*, *b*), in alloys which lie on, or very close to, a eutectic valley. This alloy therefore lies very close to the eutectic valley between the $FeAl_3$ and Co_2Al_9 fields, and use has been made of this fact in constructing figure 9. The extracted residue from alloy 1·50/2·29 confirmed the microscopical observation; the Co_2Al_9 crystals were easily distinguishable from the $FeAl_3$ crystals which, in this series of alloys, possessed a plate-like habit (figure 12, plate 9).

The remaining two alloys (1·25/2·86 and 1·00/3·45) gave primary crystals of $FeAl_3$.

The analytical results are summarized in table 2 and plotted in figure 13. Since alloy 1·50/2·29 deposited both types of crystal, the composition of the Co_2Al_9 crystals will represent the limiting solubility of iron in Co_2Al_9 under the cooling conditions employed; while the composition of the $FeAl_3$ crystals will similarly represent the limiting solubility of cobalt in $FeAl_3$.

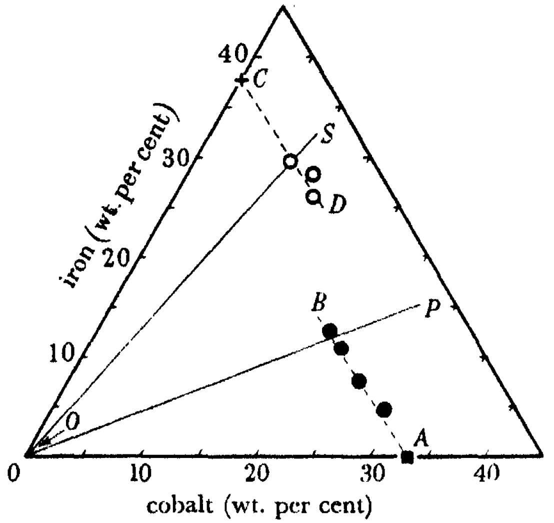


FIGURE 13. Analyses of extracted primary crystals.

TABLE 2

alloys	type of extract	composition, wt. % cobalt	composition, wt. % iron
2·25/0·57	prisms, Co_2Al_9	29·01	4·52
2·00/1·14	prisms, Co_2Al_9	25·35	7·55
1·75/1·71	prisms, Co_2Al_9	22·17	10·79
1·50/2·29	prisms, Co_2Al_9	20·18	12·47
	plates, $FeAl_3$	12·40	25·92
1·25/2·86	plates, $FeAl_3$	11·01	28·13
1·00/3·45	plates, $FeAl_3$	8·70	29·49

Although the X-ray investigations of Parker (1945) show clearly that the binary aluminium-cobalt compound is based on a ratio of 2 cobalt atoms to 9 aluminium atoms, the composition, as determined by analysis of good quality extracted crystals, indicates a cobalt content (33·2 wt. % cobalt) slightly greater than would correspond to Co_2Al_9 (Raynor & Pfeil 1946–7*b*). This composition has been used in figure 13. The line *AB* represents the replacement of cobalt by iron, atom for atom, in the binary compound of experimentally determined composition. The analyzed compositions of the crystals from ternary alloys lie very close to this line, and establish the mode of replacement. The line *OP* of figure 13 is the extrapolation of the $(\alpha + Co_2Al_9)/(\alpha + Co_2Al_9 + FeAl_3)$ phase boundary of the 640° C isothermal. The point at which it cuts the atomic replacement line represents the limiting solubility of iron in Co_2Al_9 at 640° C; this is in good agreement with the work on slowly cooled alloys.

The composition plotted for $FeAl_3$ on the aluminium-iron axis of figure 13 is that determined by Raynor & Pfeil (1946–7*a*), and the line *CD* represents atomic replacement of iron by cobalt in this compound. The analytical results for crystals extracted from the ternary alloys are in good agreement with this line. The limiting composition, given by the extract from alloy 1·50/2·29, is richer in cobalt than that indicated by the examination of solid alloys annealed at 640° C, as deduced from the intersection of the $(\alpha + Co_2Al_9 + FeAl_3)/(\alpha + FeAl_3)$ boundary (*OS*) with the atomic replacement line.

7. X-RAY EXPERIMENTS

As a final test of the correct identification of the primary crystals extracted from the alloys, X-ray diffraction patterns were obtained from both types of crystal in alloy 1·50/2·29. These corresponded very closely with the diffraction patterns of $FeAl_3$ and Co_2Al_9, and no evidence was obtained of other diffraction patterns which might be associated with a third, ternary, phase in equilibrium with the aluminium-rich solid solution.

8. DISCUSSION

The theory of the role of transitional metal solutes in aluminium-rich alloys assumes that, when alloyed with aluminium, transitional metal atoms absorb from the structure as a whole a number of electrons per atom corresponding to the available vacancies in their atomic orbitals, which for chromium, manganese, iron, cobalt and nickel are respectively 4·66, 3·66, 2·66, 1·71 and 0·61 per atom.

In discussing the constitution of aluminium-iron-nickel alloys from this point of view, it was suggested (Raynor & Pfeil 1946–7*a*) that this system could be regarded as built up from the systems aluminium-iron-cobalt and aluminium-cobalt-nickel, placed side by side. This implies that, from a knowledge of the aluminium-nickel-iron system, it is possible to predict the general forms of the other two ternary diagrams. It has been shown that, as regards the aluminium-cobalt-nickel alloys, the predictions are remarkably well satisfied (Raynor & Pfeil 1946–7*b*). It is now of interest to compare the results of the present work with the predictions made in connexion with the aluminium-iron-cobalt system.

In the first place, it is clear from all the experimental evidence that, as suggested, no ternary compound exists in equilibrium with the primary solid solution. Just as equilibrium in the aluminium-iron-nickel alloys which contain an excess of iron over that required for an Fe:Ni ratio of 1 : 1 involves only the solid solution, $FeAl_3$ and $FeNiAl_9$, so in the present system only the solid solution, $FeAl_3$ and Co_2Al_9 come into equilibrium with each other.

Secondly, in agreement with expectations, the compound $FeAl_3$ will dissolve appreciable amounts of cobalt. The compound Co_2Al_9, however, which was expected, by comparison with $FeNiAl_9$, to dissolve only limited amounts of iron, proves to have a relatively large solubility for this element.

With the exception of the solubility of iron in Co_2Al_9, to which further reference is made below, the general form of the constitutional diagram is in good agreement with the theoretical expectations.

The results obtained for the previously unknown systems aluminium-iron-cobalt and aluminium-cobalt-nickel make it possible to show how closely the original suggestion, that the aluminium-iron-nickel diagram may be regarded as built up from those for the other two systems, is in agreement with the facts. In figure 14, we have plotted the 600° C data from the two simpler systems in such a manner that each diagram occupies a triangle with angles of 90, 60 and 30°; atomic percentages have been used. Thus, the triangle *ABC* represents the 600° C isothermal of the aluminium-iron-cobalt alloys, while the triangle *ACD* represents the corresponding diagram for the aluminium-cobalt-nickel alloys. The two triangles placed together, with the aluminium-cobalt axes coincident, make up the usual equilateral triangle employed for plotting ternary systems. Figure 15 shows the corresponding isothermal for the aluminium-iron-nickel alloys, and the general correspondence between figures 14 and 15 is at once apparent.

It will be appreciated that this method of plotting is equivalent to plotting the two figures on the same scale of electron : atom ratios; the variations of the electron : atom ratio along any line of constant aluminium content is practically linear between the same terminal values in each case.

It is particularly to be noted that the correspondence of the aluminium-cobalt-nickel triangle in figure 15 with the relevant portion of the aluminium-iron-nickel diagram is almost exact, while, at this temperature, the solubilities of cobalt and of nickel in $FeAl_3$ along the $(\alpha + FeAl_3)/(FeAl_3)$ boundary proceed to closely similar electron : atom ratios.

Several further points of interest arise from this work. The compounds Co_2Al_9 and $FeNiAl_9$ are isomorphous, and it has been previously shown that nickel will replace cobalt in the former and iron in the latter in such a manner that the replacement proceeds atom for atom (Raynor & Pfeil 1946–7*b*). Further, the replacement continues until, in each case, an electron : atom ratio of 2·285 is reached, calculated on the basis of the theory referred to above. It is probable, therefore, that the Brillouin zone for the Co_2Al_9 type of structure can tolerate this number of electrons per atom before becoming unstable with regard to possible alternative structures. In the same way, $FeAl_3$ dissolves cobalt and nickel respectively to closely similar electron : atom ratios. This suggests that structures of this type are to be regarded as essentially based on the electron : atom ratio.

According to the present work (§6), Co_2Al_9 (electron : atom ratio 2·14) will dissolve iron in such a manner that cobalt atoms are replaced by iron atoms until an electron : atom ratio of 2·07 is reached. The deficit, 0·07 electron per atom, is less than half the surplus which can be accommodated when nickel is dissolved in this structure. Similar indications are given by a consideration of solid solution formation in other binary compounds of aluminium with transitional elements. $MnAl_6$ will dissolve a large amount of iron, the replacement occurring atom for atom until approximately half the manganese is replaced, whereas only a limited solubility of chromium exists. Similarly, $FeAl_3$ dissolves appreciable quantities of cobalt, but little manganese, the replacement again being atom for atom in each case. $NiAl_3$, however, does not dissolve any of the transitional elements preceding nickel in the periodic table to any appreciable extent. There is thus a greater tendency to solid solution formation on increasing the number of electrons per atom than on decreasing this value. These facts may be understood in terms of the ability of the Brillouin

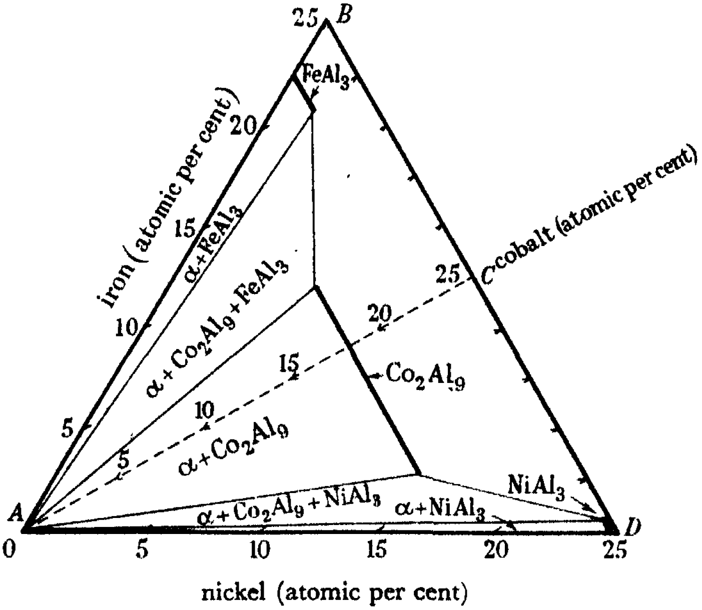


FIGURE 14

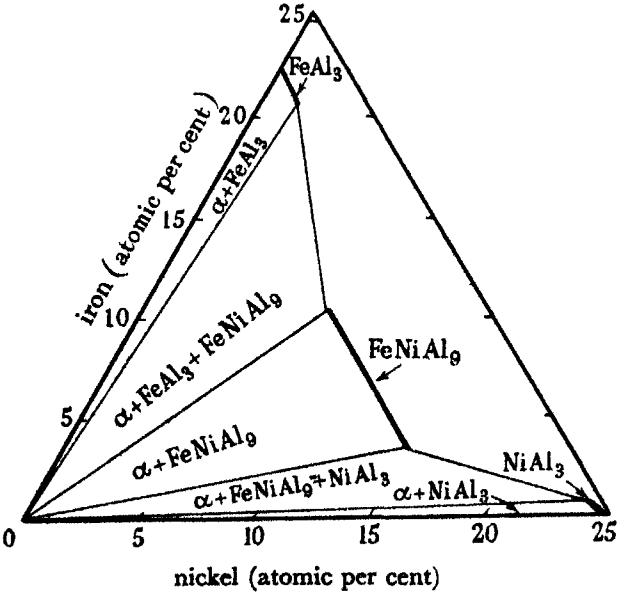


FIGURE 15

zones for these compounds to accommodate extra electrons, and the inability of the structure to hold together when too few electrons are present.

Although the Co_2Al_9 structure can tolerate a certain deficit of electrons, $FeNiAl_9$ cannot tolerate a corresponding deficit, and dissolves very little iron. Though the effects of increasing the number of electrons per atom depend upon the Brillouin zone characteristics, the behaviour on decreasing the number of electrons per atom probably depends on the details of the atomic arrangement in the structure itself. In this connexion, it should be noted that, whereas Co_2Al_9 is stable up to a temperature of 943° C, $FeNiAl_9$ is formed, in aluminium-iron-nickel alloys, by reaction between $FeAl_3$ and $NiAl_3$ at the considerably lower temperature of 809° C. This indicates that the structure in which half the transitional metal atom sites are occupied by iron, and half by nickel, is less stable than that in which cobalt atoms occupy all the relevant sites. The stability of $FeNiAl_9$ might therefore be expected to be more sensitive than that of Co_2Al_9 towards a reduction in the number of valency electrons per atom, which in general tends to reduce the cohesive forces in a metallic structure; this may be responsible for the ability of iron to dissolve in Co_2Al_9, but not to any appreciable extent in $FeNiAl_9$.

This research was carried out in the Metallurgy Department of the University of Birmingham, under the general supervision of Professor D. Hanson, D.Sc., to whom the authors' thanks are due for his interest and support. The authors must also acknowledge the valuable assistance of Mr G. Welsh in the experimental work, and the analytical work of Mr P. Wingrove.

The authors express their gratitude to the Department of Scientific and Industrial Research, The Royal Society, the Chemical Society, and Imperial Chemical Industries, Ltd., for general financial assistance.

REFERENCES

Archer, R. S. & Fink, W. L. 1928 *J. Inst. Metals*, **40**, 356.
Bradley, A. J. & Taylor, A. 1938 *Proc. Roy. Soc.* A, **166**, 353.
Bradley A. J. & Taylor, A. 1940 *J. Inst. Metals*, **66**, 53.
Dix, E. H. (Jnr.) 1925 *Proc. Amer. Soc. Test. Mater.* **25** (II), 120.
Fink, W. L. & Freche, H. R. 1932 *Trans. Amer. Inst. Min. (Metall.) Engrs.* **99** (Inst. Metals Div.), 141.
Parker, Audrey M. B. 1945 *Nature*, **156**, 783.
Pauling, L. 1938 *Phys. Rev.* (ii), **54**, 899.
Raynor, G. V. 1944 *J. Inst. Metals*, **70**, 531.
Raynor, G. V. 1945 *Phil. Mag.* **36**, 770.
Raynor, G. V. & Little, K. 1945 *J. Inst. Metals*, **71**, 493.
Raynor, G. V. & Pfeil, P. C. L. 1946–7*a* *J. Inst. Metals*, **73**, 387.
Raynor, G. V. & Pfeil, P. C. L. 1946–7*b* *J. Inst. Metals*, **73**, 609.

A comparative X-ray study of foetal and adult sheep haemoglobins

By J. C. Kendrew and M. F. Perutz

Cavendish Laboratory and Molteno Institute, University of Cambridge

(*Communicated by D. Keilin, F.R.S.—Received* 22 *January* 1948)

[Plate 10]

Haemoglobins from adult sheep and from foetuses of different ages were crystallized, using a standard procedure. The optical properties, cell dimensions and space-groups of the wet and, where possible, of the air-dried crystals were determined, and surveys were made of the lower order diffraction patterns in each of the principal crystal zones.

The results obtained provide additional evidence for the non-identity of the foetal and adult haemoglobins and lead to certain conclusions relevant to the general problem of haemoglobin structure.

Methaemoglobin from foetuses aged less than 120 days was found to crystallize in either of two forms, each of which was different from methaemoglobin crystals of adult sheep prepared by the same method. Haemoglobin from later foetuses is difficult to crystallize, and its properties resemble those of artificial mixtures of adult and early foetal haemoglobins.

No general picture of the structural differences between adult and foetal haemoglobin has emerged from this study, mainly because these were obscured by differences in molecular packing. Indications were found of a difference in molecular symmetry between these two proteins, and this could be correlated with different splitting properties in solution. There is also evidence of a pronounced intramolecular layering in foetal haemoglobin.

A new method of controlled shrinkage of protein crystals at constant humidity was developed and applied to foetal haemoglobin crystals, which were shown to pass through a series of at least four well-defined lattices.

The densities of wet adult and foetal haemoglobin crystals were measured and used for the estimation of 'bound water'. The amounts of 'bound water' were of the same order as those obtained from haemoglobin crystals of other species.

1. INTRODUCTION

The relation between the haemoglobins obtained from adults and foetuses of the same animal species has been a subject of interest to physiologists for many years, and many physical and chemical techniques have been applied to its investigation. Interest in the subject began with the observation that foetal and adult bloods differed in their most characteristic physiological property, namely, the oxygen dissociation curve. Generally, the foetal curve lies to the left of the adult one when plotted according to the usual convention; that is to say at any given oxygen pressure foetal haemoglobin is oxygenated to a greater extent than adult. In addition, the two curves differ in shape; in adults the curve is pronouncedly sigmoid while in foetuses it approaches a rectangular hyperbola. There are numerous complications—e.g. the effects of concentration and of pH, and the differences in behaviour between haemoglobin in the corpuscle and in solution—nevertheless, it seems clear enough that the dissociation curves are intrinsically different in all the species examined. These differences have been studied by many workers over a long period; the present position has been summarized by Barcroft (1946).

These observations stimulated a great many attempts to find other physical differences between adult and foetal haemoglobins, attémpts which were generally successful. A brief enumeration of the more important of the differences reported will sufficiently indicate their variety; rate of alkaline denaturation (man: von Krüger 1925; von Krüger & Bischoff 1925; von Krüger, Bischoff & Schulte 1926; Haurowitz 1930; Brinkman, Wildschut & Wittermans 1934; Brinkman & Jonxis 1935), rate of spreading in monolayers at the isoelectric point (man, goat, cow: Brinkman & Jonxis 1935, 1936), crystalline form (man: Nicoletti 1930, 1932; Perrier & Jannelli 1931; Haurowitz 1935), immunological specificity (man: Darrow, Nowakowsky & Austin 1940), solubility (cow: Wyman, Rafferty & Ingalls 1944), electrophoretic mobility (man: Andersch, Wilson & Menten 1944; sheep: Adair 1939); number of terminal amino-acid residues (man: Porter & Sanger 1948).

Fundamentally, the discussion of the problem of the non-identity of foetal and adult haemoglobin involves an acceptance of the idea that the term 'chemical individual' can properly be applied to proteins; granted this—and it would now seem to be established beyond all doubt—*any* difference in property between two proteins, when observed under identical environmental conditions, argues conclusively a true 'chemical' or structural difference between them.

Hence the unanimous verdict of so diverse a series of physico-chemical tests leaves little room for doubt that adult and foetal haemoglobins from the same species do in fact possess different molecular structures. Nevertheless, it was thought by the present authors that the methods of X-ray crystallography could usefully be applied to the problem, because potentially they are capable not only of establishing whether two proteins are structurally dissimilar, but also of revealing the nature of some of the structural differences between them. The techniques of crystallography and X-ray analysis are particularly applicable to the present investigation, since the haemoglobins readily form crystals of unusual perfection. Fortunately, too, identical methods of crystallization and examination can be used, thus satisfying the above-mentioned criterion of identical environments, which is, of course, crucial and which the peculiar characteristics of proteins sometimes make it difficult to satisfy.

Scope of the present research

Recently a detailed investigation has been carried out by one of us, with various collaborators, into the crystal stucture of horse methaemoglobin (Boyes-Watson, Davidson & Perutz 1947), and, in addition, an X-ray study has been made of human haemoglobin (Perutz & Weisz 1947). The techniques developed in these studies are suitable for crystalline proteins in general, and have been used in the present researches on foetal and adult haemoglobin. The species chosen was the sheep because of the ready availability of samples of foetal and adult sheep blood, provided by the generosity of the late Sir Joseph Barcroft, at whose suggestion the research was carried out.

The study described in this paper was begun with two objects in mind. First, it was hoped to add yet another type of evidence to those enumerated above concerning the differences between adult and foetal haemoglobin and, if possible, to discover the nature of at least some of the structural differences between them. Secondly, the choice of sheep as a species enabled us to extend the comparative studies of the haemoglobins, the species specificity of which has been well known since the classical morphological studies made by Reichert & Brown (1909) on haemoglobins crystallized from numerous animal species. This species specificity, which occurs in many types of protein, must have a basis in differences of molecular structure, but the nature of these differences has yet to be found.

Unfortunately, the complete realization of these objects is beset by certain difficulties which are inherent in the X-ray analysis of substances as complex as the crystalline proteins. In the first instance it must be admitted that nothing short of a complete structure analysis can, except in favourable circumstances, provide incontrovertible evidence of a definite molecular difference between two substances so closely related as foetal and adult haemoblobin seem likely to be. Of course the application of X-ray methods makes it possible to dispose with certainty of the case where two apparently dissimilar crystals are different only in habit but identical in internal structure, containing the same molecules arranged in the same way in the same unit cell and space-group, and therefore produce identical diffraction patterns. But quite apart from this the frequently occurring phenomenon of polymorphism makes it possible for the same molecular species to form two or more types of crystal which are different not only morphologically but also in respect of the X-ray patterns they produce, the mode of packing of the molecules into a crystal lattice being different in the different cases. For instance, we were able to prepare two distinct crystalline materials from foetal haemoglobin. These crystals belong to different space-groups, have different cell dimensions—one cell containing four molecules and the other two—and yield diffraction patterns dissimilar in every respect. Yet, as will be shown in the sequel, there is little doubt of the identity of the proteins in the two materials. This example underlines the danger of drawing positive conclusions from X-ray evidence in the absence of a complete structure determination and illustrates the point that differences of diffraction pattern can only establish a prima facie case for molecular differences, a case which must be supported by other evidence which is in general not fully conclusive.

As to our second object, possibilities of discovering the structural basis of species specificity will depend, at the present stage of X-ray technique, on the nature of the structural differences involved. They may be on a gross scale, involving changes in the size of the asymmetric unit,* in the relative positions of subunits and in the arrangement of the polypeptide chains within them, or, alternatively, there may only be changes in the detailed composition, amino-acid sequence, and stereochemical configuration within the chains, the general arrangement remaining unaltered. Experience gained in the X-ray analysis of horse haemoglobin suggests that the former type of specificity should be detectable, while the latter probably lies beyond the powers of resolution of our present methods even if the analysis were pushed to the limit. Not even gross structural characteristics (apart from the size of the asymmetric unit) can at present be elucidated, however, without a very detailed and lengthy analysis to which different types of crystal do not lend themselves equally readily. Through accidents of habit and molecular packing some forms of protein crystal are capable of yielding much more information than others, and until the more superficial comparative studies have surveyed the field it is not possible to determine where a more intensive attack would be most profitable. As will be indicated below, none of the crystals of foetal or adult sheep haemoglobin proved to be particularly suitable for detailed analysis; the present study, therefore, was more concerned with a general survey of the types of crystal structure which occur in the foetal-adult haemoglobin system of sheep, than with a far-reaching examination of any one of them, and it represents no more than a beginning in the field of comparative studies of the crystalline proteins.

---

* The 'asymmetric unit' is a term frequently used in X-ray analysis to define the smallest unit of pattern in the structure. If V is the unit cell volume, d the crystal density, s the number of general (asymmetric) positions in the unit cell as determined by the space-group symmetry, and N Avogadro's number, then the molecular weight of the asymmetric unit $M_a = VNd/s$. If the molecular weight M of the substance is known, M/M_a may be an integer (p), in which case the molecule consists of at least p chemically and structurally identical subunits.

An account is first given of the techniques employed, both preparative and X-ray; and this is followed by a description of the morphological and optical properties of the crystals obtained, and of the results bearing on the foetal-adult haemoglobin problem which were obtained by study of their X-ray diffraction patterns. Finally, we enumerate a number of findings, relevant to general questions of protein structure, which emerged in the course of the research.

2. EXPERIMENTAL TECHNIQUES

Crystallization of haemoglobins

The methods used were essentially those described by Keilin & Hartree (1935), as modified by Perutz (see Boyes-Watson *et al.* 1947). In all cases the haemoglobin was oxidized to methaemoglobin prior to crystallization, since methaemoglobins can be stored without special precautions; it transpired that crystals of methaemoglobin could be kept in their mother liquor for periods up to 6 months at room temperature without apparent change. Recent work on horse haemoglobin has shown that there is no essential difference between the structures of the oxy-, carboxy- and met- forms of the same haemoglobin, the crystals of all three being isomorphous (in contrast to reduced haemoglobin which possesses an entirely different crystal structure).

In brief, the method was to remove the serum from whole blood by repeated washing with isotonic saline, followed by haemolysis of the corpuscles by simple freezing at − 20° C. Stromata were then removed by filtering through celite. The haemoglobin was oxidized by adding saturated potassium ferricyanide. Potassium ferrocyanide and excess ferricyanide with other impurities were then removed by dialysis against water. Finally, crystallization was effected by salting out with ammonium sulphate, the more insoluble proteins being removed by a preliminary half-saturation. Simple dialysis cells with cellophane membranes were employed, and in this way it was possible to carry out slow crystallization over periods of about a week, with production of large crystals. In general, it was necessary to control the pH of the solution by means of phosphate buffers, a series of dialysis cells at different pH's being set up for each protein in order to find the optimum. In practice a relatively small change of the order of 1 to 2 pH units is sufficient to alter completely the habit and size of the crystals formed.

In the case of adult sheep methaemoglobin it was found that precipitation began at about 70 % saturation with ammonium sulphate, and that the most satisfactory crystals were obtained at pH's between 5 and 6, unbuffered ammonium sulphate being suitable in this instance.

The sample of foetal sheep's blood used for the main part of the present researches was supplied by Sir Joseph Barcroft, being derived from a foetus aged approximately 95 days (sheep number 685*); it may be noted that sheep are normally born on the 147th day of pregnancy. The foetal haemoglobin was found to be precipitated at between 60 and 65 % saturation with ammonium sulphate; careful parallel tests with blood samples from the mother of the foetus, worked up under identical conditions, showed that the difference in solubility was quite real. A similar difference nas been noted by Karvonen (1947). Optimum conditions for crystallization of foetal haemoglobin were found to be at a pH of 6·5, and as will be indicated later, on one occasion a mixture of two forms was obtained in these circumstances, though in all other cases only one type of crystal was formed.

Technique of X-ray analysis

The methods adopted in mounting wet crystals and in photographing their X-ray diffraction patterns were in general closely similar to those employed in the work on horse haemoglobin (Boyes-Watson *et al.* 1947) and will not be further described here. All photographs were taken with Cu Kα-radiation on flat plates, the specimen-plate distance being 6 or 10 cm.

One change in technique was the use of a 'hanging backstop' to reduce air-scattering in the central regions of the photographs.† Instead of the conventional method of mounting the backstop immediately in front of the film, a lead disk was suspended by means of a glass fibre from a head fitted with translation movements, the disk being about 1 cm. in front of the plate. Adjustment was carried out optically. By this simple method photographs were obtained which, while not so free from air-scattering as those taken with a vacuum camera, were quite adequate in most cases. The distance of 1 cm. between backstop and plate was arrived at by trial as a compromise; at larger distances the scattering is reduced even further, but as the distance increases the backstop cuts off progressively more of the lower angle reflexions.

3. OPTICAL AND X-RAY EXAMINATION OF CRYSTALLINE ADULT AND FOETAL SHEEP METHAEMOGLOBINS

We now proceed to an account of the morphological and optical properties of the crystals of adult haemoglobin and of two forms of foetal haemoglobin, and of their X-ray analysis in so far as it is relevant to the foetal-adult haemoglobin problem.

---

* Sheep numbers mentioned in this paper are those allocated by Barcroft and referred to in his publications.

† This technique was also used by Crowfoot (1938) in her studies of crystalline insulin.

We do not propose to discuss the general properties of haemoglobin crystals (e.g. shrinkage on drying) which dictate the use of special techniques of analysis, since these properties and techniques have been fully described elsewhere. All the haemoglobin crystals examined here, both adult and foetal, give the type of diffraction photograph which has now become familiar for protein crystals. In the wet state sharp reflexions extend to spacings of about 2·5 A; on drying the crystals this pattern disappears to a greater or lesser extent, and only comparatively few reflexions around the centre of the photograph remain.

The best value for the molecular weight of adult sheep haemoglobin is 68,000 (Adair 1928). The same figure is taken for foetal sheep haemoglobin in accordance with the findings of McCarthy & Popják (1943), who state that there is no significant difference between the values for foetal and adult sheep haemoglobins, as determined by osmotic pressure measurements. Determinations of iron content indicate that four haem groups are present per molecule in each case.

For convenience a comparative table of the cell dimensions of all forms of sheep haemoglobin investigated has been drawn up in table 1.

TABLE 1. UNIT CELL DIMENSIONS, ETC., OF ADULT AND FOETAL SHEEP METHAEMOGLOBINS

		orthorhombic foetal			monoclinic foetal		
	adult wet	wet	dry	% contraction	wet	dry	% contraction
space-group ...	$C2$	$B22_12$			$P2_1$		
number of mol./cell ...	4	4			2		
a	164 A	112 A	78 A	30	55 A	48 A	13
b	70 A	108 A	99 A	8	83 A	71 A	14
c	66 A	56 A	54 A	4	64 A	58 A	9
β	94·5°	—	—	—	102°	101°	—
cell volume	755,000 A^3	677,000 A^3	417,000 A^3	38	286,000 A^3	194,000 A^3	32

(i) ADULT SHEEP METHAEMOGLOBIN

Morphological and optical properties

Adult haemoglobin was found to crystallize in only one form, and this is shown in figures 1 and 2, plate 10, to consist characteristically of triangular plate-like crystals. The plates were found to be monoclinic with the a and b axes lying in the plane of the plate and the a axis usually being parallel to one of the long edges. The faces around the edges of the plate are bevelled and correspond to the faces of a distorted rhombohedron (figure 2). Determination of the direction of the three optic axes was not practicable, since the crystals were generally too thin for cutting sections normal to the (001) face.

The crystals are strongly pleochroic when viewed with the light incident normal to (001), the direction of greatest absorption being parallel to [010], as in monoclinic horse haemoglobin. It may be supposed that the strong pleochroic absorption is due to the conjugated ring systems of the haem groups in the molecule, all four haem groups being more or less parallel and reinforcing one another optically. On this hypothesis the b axis lies in the plane of the haem groups, which would be approximately normal to (001).

Cell dimensions and space group

Oscillation photographs of the wet crystals, taken about the three principal crystal axes, gave the unit cell dimensions $a = 164\,\text{A}$, $b = 70\,\text{A}$, $c = 66\,\text{A}$, and $\beta = 94{\cdot}5^\circ$. Reflexions are absent for $(h+k)$ odd; no other systematic absences were observed. Hence the possible space-groups are $C2$, Cm and $C2/m$. Of these the two latter are excluded, since crystals containing only one stereoisomer of an optically active compound cannot belong to any of the non-enantiomorphous space-groups. The space-group must therefore be $C2$, which has four general positions in the unit cell; the number of molecules per cell must be a multiple or submultiple of this number.

The density of the wet crystals suspended in 3·25 M-ammonium sulphate solution was measured by the method of McMeekin & Warner (1942), as described by Perutz (1946), and found to be 1·225. The crystals shrink visibly on drying, indicating that when wet they contain a considerable proportion of liquid of crystallization. Unfortunately the diffraction pattern obtained from the dry crystals was so poor that it was not even possible to measure their cell dimensions. Hence the number of haemoglobin molecules per unit cell, which is normally derived from the cell volume, density and water content of the *dried* crystals has in this case to be calculated from the experimental data for the *wet* crystals. This can be done in several ways, each of which involves the introduction of an arbitrary but plausible assumption.

For example, from the density and from the wet-cell volume we may calculate the cell weight as 561,000 in molecular weight units, i.e. just over 8 times the molecular weight of 68,000. The degree of shrinkage observed on drying makes it highly unlikely that the number of molecules per cell is as great as 8 (which would leave room for only 3 % of water). The far more likely number is 4—i.e. the same as the number of general positions—an assumption which leads to a value of 55 % for the percentage of liquid of crystallization in the wet cell, corresponding closely to the figure of 52 % for horse haemoglobin. Smaller numbers of molecules per cell (2 or 1) would give much higher percentages of liquid of crystallization than have been observed in any other protein.

The number of molecules per unit cell can also be found by making the alternative assumption that, as in the case of horse haemoglobin, the concentration of ammonium sulphate in the liquid of crystallization is about two-thirds that in the suspension medium. This gives a value for the specific volume of the liquid of crystallization; the specific volume of the protein is known (Adair & Adair 1936). From these figures we may calculate the weight fraction of protein in the system, and hence deduce the number of molecules in the unit cell. This calculation leads to the value 3·9.

We may therefore take it as very probable that the number of molecules per unit cell is 4, i.e. the same as the number of general positions. This means that no intrinsic symmetry elements can be deduced for the molecule, though there is equally no evidence for the absence of such elements. There are no striking absences or powerful reflexions among the low orders of $h00$, $0k0$, or $00l$, so it is not possible to draw any immediate deductions about the packing of the molecules in the cell.

(ii) FOETAL SHEEP METHAEMOGLOBIN

Mention has already been made of the fact that foetal haemoglobin can crystallize in two forms. One of these, the needle-shaped or orthorhombic type, is invariably obtained on salting out with ammonium sulphate under suitable conditions; the other, known as the monoclinic type, was obtained on one single occasion only in admixture with the orthorhombic type, and is presumably less stable. The two forms will be described in turn.

(*a*) *Orthorhombic form: morphological and optical properties*

This, the common form of foetal sheep haemoglobin, appears as thin needles as shown in figure 3, plate 10. The needles sometimes show terminal dome faces, and another characteristic phenomenon is that of a re-entrant angle at one end of the needle—due presumably to twinning. In some preparations there is a tendency for the needles to flatten out into long laths of extreme thinness (as in the photograph).

The crystals exhibit straight extinction, the slow ray coinciding with the needle axis. They are strongly pleochroic, the direction of the electric vector for maximum absorption also coinciding with the needle axis, which was chosen as b by analogy with the adult crystals. As in the latter case the deduction may be made that the haem groups are parallel within each molecule, and that the molecules are so disposed in the crystal that the haem groups are parallel throughout the crystal, in planes containing the needle axis and roughly normal to the flat face of the needle.

X-ray examination

The study of these crystals presented considerable difficulties on account of their unfavourable habit. To obtain even a feeble diffraction pattern long exposures of 12 to 24 hr. were required, and it was also necessary to use crystals which were unusually large in two dimensions, in order to compensate for their excessive thinness in the third, and this meant that spots adjacent in the pattern tended to fuse together, making indexing difficult. This trouble was aggravated by the fact that two of the crystallographic axes turned out to be very long (over 100 A).

The unit-cell dimensions of the wet crystals are $a = 112\,\text{A}$, $b = 108\,\text{A}$ and $c = 56\,\text{A}$; the orientation of the three axes can be seen from the drawing in figure 4, plate 10. It will be noted that a and b are nearly equal and that c is roughly half of each of them, relations which probably explain the occurrence of various types of twin. Absences are hkl when $(h+l)$ is odd and $0k0$ when k is odd. Ruling out non-enantiomorphous space-groups for the reason explained above, $B22_12$ is the only one compatible with the X-ray data. This has eight general positions in the unit cell.

On drying in air the crystals shrink considerably and break up very badly. 5° oscillation photographs of the dried crystals showed only two or three spots even when long exposure times were used, and to interpret them it was necessary to assume that the space-group had remained the same during shrinkage (an assumption found to be justified in the cases of all protein crystals hitherto examined). On this basis the dry-cell dimensions are $a = 78$ A, $b = 99$ A and $c = 54$ A; it may be seen from table 1 that the a dimension has shrunk to a very much greater extent than the other two.

The crystals had to be dried from ammonium sulphate solution which always left a certain amount of salt included within the crystals, and hence their density could not be measured. Assuming a value of 1·27 for the dry density, by analogy with horse methaemoglobin and other protein crystals, the cell weight is 324,000 in molecular weight units which clearly sets an upper limit of four for the number of molecules of weight 68,000 per unit cell. Since there are eight general positions per unit cell, this means that each molecule must lie in a special position on a diad axis, and must therefore consist of two structurally identical halves.

(*b*) *Monoclinic form: morphological and optical properties*

These crystals form parallelogram-shaped plates, far thicker than the orthorhombic needles described above, and with edges bevelled in the manner of a distorted rhombohedron. For light passing through the crystal normal to the largest face pleochroism could not be detected; for light parallel to the flat face only slight pleochroism was observed. Extinction was oblique, the slow ray being parallel to the short diagonal of the principal crystal face. It is clear that in this type of crystal the haem groups do not all lie in parallel planes throughout the structure; in fact they must be disposed so that their optical anisotropies practically cancel one another in all directions. Unfortunately, the few crystals available were used up before their morphology and optics had been fully determined.

Unit-cell dimensions and space-group

The cell dimensions of the wet crystals are $a = 55$ A, $b = 83$ A, $c = 64$ A and $\beta = 102°$. The only systematic absences are $0k0$ for k odd. Excluding once again non-enantiomorphous space-groups, $P2_1$ is the only one compatible with the X-ray data. This contains two general positions in the unit cell.

On drying in air the crystals retain as high a degree of order as do orthorhombic horse methaemoglobin or insulin; faint spots can still be seen in the diffraction pattern at positions corresponding to spacings of 7 or 8 A. This small degree of disordering may perhaps be correlated with the low liquid content of the crystals in the wet state—only 41 % by volume, as compared with 55 % for adult sheep haemoglobin and 52·4 % for monoclinic horse haemoglobin.

Measurement of the density of the dry crystals was again impracticable, since the crystals had been dried from ammonium sulphate solution. However, assuming as before that the density of the dried crystals is 1·27, the weight of the dry unit cell is 148,500 in molecular weight units, which is incompatible with any number other than two molecules of weight 68,000 per cell. It should perhaps be pointed out that the assumption about the density of the dry crystals is not so arbitrary as it may seem, since all proteins have practically identical partial specific volumes in solution, and the density of air-dried crystals generally seems to vary only between about 1·25 and 1·29 in accordance with individual differences in molecular packing and water content.

The space-group of the dried crystals is the same as that of the wet, viz. $P2_1$. The dry dimensions are given in column 7 of table 1. It will be seen that the shrinkage effects are quite different from those of horse methaemoglobin, which also possesses a monoclinic unit cell. In the latter case a remained constant until the last stage of drying, while b and c remained almost constant throughout; on the other hand, β changed from 111 to 137·5°. In other words, in horse haemoglobin shrinkage takes place almost entirely by a shearing of the whole cell, the value of $c \sin \beta$ altering from 51·4 to 36·5 A; and this has been interpreted as due to a layered arrangement of the molecules in the lattice, the layers approaching one another more closely as water is removed from between them. By contrast, in foetal sheep haemoglobin we find that β remains practically constant while the three cell dimensions all shrink. The final percentage contraction in the cell volume is the same, viz. 32 %, but we may deduce that the disposition of the molecules and of the liquid of crystallization must be entirely unlike that in horse haemoglobin.

4. ON THE NON-IDENTITY OF FOETAL AND ADULT SHEEP HAEMOGLOBINS

The foregoing section has shown that in point of crystal structure, adult sheep haemoglobin differs entirely from both the forms of foetal haemoglobin which were examined. The differences extend to all characteristics of the structure—space-group, cell dimensions, water content, mode of shrinkage and the details of the diffraction patterns. But it will also have been noted that the differences between the two types of crystal obtained from foetal haemoglobin solutions are quite as great as those between adult crystals and either form of foetal crystal. Further evidence must therefore be adduced in support of any contention that adult and foetal haemoglobins are distinct molecular species, since the crystallographic evidence alone for such a distinction could be paralleled by equally good evidence that monoclinic and orthorhombic foetal crystals contained different molecular species.

The identity of monoclinic and orthorhombic foetal haemoglobin

The last possibility can readily be disposed of. In the preparation containing monoclinic foetal crystals there was, as already mentioned, a mixture of this variety with the normal orthorhombic ones. It was, however, possible to isolate with a needle a number of monoclinic crystals, and to dissolve these in distilled water. From the resulting haemoglobin solution a fresh crop of crystals was prepared by salting out, and examination showed this new crop to be identical with the needle-shaped, orthorhombic type. Although it was not possible to carry out the reverse transformation (as already stated, the monoclinic type of crystal was only obtained on a single occasion), there seems to be no doubt that the two kinds of crystal are in fact modifications of the same protein.

Recrystallization experiments with adult and orthorhombic foetal haemoglobins

A totally different behaviour on recrystallization was exhibited by specimens of adult haemoglobin and of orthorhombic foetal haemoglobin. In every experiment with either material the recrystallized product was morphologically, optically and (in the few cases tried) in regard to its X-ray diffraction pattern, identical with the parent preparation. That is to say, specimens of adult haemoglobin invariably recrystallized to monoclinic plates of the type described earlier in this paper, while foetal specimens formed characteristic needles of orthorhombic type. The recrystallized product appeared to contain one type of crystal only, and to be free from amorphous material. It should be emphasized that in these recrystallization experiments the conditions of pH, protein concentration, etc., were the same in parallel tests carried out on adult and foetal specimens, the only difference being that foetal haemoglobin was normally precipitated at 65 % and adult at 70 % saturation with ammonium sulphate.

Behaviour of artificial mixtures of adult and foetal haemoglobin

To test the possibility that adult and foetal haemoglobins might form mixed crystals—a possibility which, if found to occur, would gravely complicate the situation—a series of artificial mixtures of adult and foetal methaemoglobin solutions was prepared, and subjected to the normal technique for crystallization. The results obtained are summarized in table 2, where *A* refers to the normal adult type of crystal, and *F* to the orthorhombic foetal variety.

TABLE 2. CRYSTALLIZATION OF ARTIFICIAL MIXTURES OF FOETAL AND ADULT SHEEP HAEMOGLOBIN

approximate ratio (adult/foetal)	form of precipitate
pure adult	*A* crystals only, no amorphous
5/1	mostly *A* crystals, some amorphous
3/2	all amorphous, no crystals
1/1	mostly amorphous, a few *A* and also a few *F*
2/3	mostly amorphous, a few *F*
pure foetal	*F* crystals only, no amorphous

These results show that adult and foetal haemoglobin have no tendency, at any rate under these conditions, to form mixed crystals. They also provide support for the hypothesis that the proteins concerned are distinct, since if the proteins were identical and the distinction between 'adult' and 'foetal' crystals were merely a case of polymorphism, such distinction would be lost in solution and artificial mixtures would be expected to precipitate entirely as crystals of one type or the other.

Experiments on haemoglobins derived from late foetuses

If, indeed, foetal blood contains a haemoglobin different from that of the adult, then it is clear that at some stage in the development of the animal there must be a change-over from the foetal to the adult form, and the presumption is that such a change-over would be more or less gradual, and that an examination of the blood obtained from the animal at this stage would reveal the presence of a mixture of the two kinds of haemoglobin. Other workers had tentatively concluded that this change-over takes place during the last month or so of foetal life, i.e. after about the 120th day of pregnancy. This view has been supported by the work of Karvonen (1947, personal communication), who plotted the solubility curves in ammonium sulphate solutions of foetal haemoglobins of various ages, using as test material samples of blood from the same sheep foetuses as were examined by us from the crystallographic point of view. Karvonen concludes from the shapes of these curves that there are in fact at least two distinct types of sheep haemoglobin, and that up to about the 120th day foetal blood contains almost exclusively the foetal form, but that after this point the proportion of adult haemoglobin increases until at birth it is present to the extent of perhaps 30 or 50 %. He furthermore finds the phenomena occurring in very young lambs' blood too complicated to be explained except on the assumption that there is at least one more distinct form of haemoglobin, making three in all (a possibility mentioned by Barcroft (1935)).

Although Karvonen worked with carboxyhaemoglobin instead of methaemoglobin, the broad outlines of his findings with young foetuses and with adult sheep corresponded closely, from the crystallographic point of view, with the results described in the earlier part of this paper, the crystalline forms of his precipitates of carboxyhaemoglobin being similar to those which we obtained with methaemoglobin. However, in the case of salted-out precipitates derived from the blood of foetuses in the last month of foetal life, new and complicated phenomena were observed. From a 144-day foetus, for example, both triangular plates and needle-shaped crystals were obtained on different occasions; from other samples he obtained a new type of crystal altogether, apparently of the same variety as those described by us below. In the case of young lambs crystalline aggregates of another new form were obtained, consisting of oval crystalline plates showing strong birefringence. In general, these haemoglobin solutions from late foetuses and young lambs were difficult to crystallize at all, much of the haemoglobin being precipitated in an amorphous form, behaviour in marked contrast to that of adult haemoglobin and of haemoglobin derived from younger foetuses, both of which crystallize completely and with great readiness.

In parallel with these experiments, we have attempted to crystallize by dialysis a number of methaemoglobin samples from late foetuses, with the following results.

(*a*) *Foetus aged* 95 *days* (*sheep no.* 685)

Samples from this foetus were used for all the work on foetal haemoglobin described earlier in this paper, and as already stated they yielded orthorhombic needle-shaped crystals, crystallization being complete and the crystals being all alike.

(*b*) *Foetus aged* 120 *days* (*sheep no.* 682)

This foetus gave results exactly similar to those of (*a*).

(*c*) *Foetus aged* 129 *days* (*sheep no.* 689)

This time crystallization was very slow, but finally there appeared a small crop of needle-shaped crystals which turned out to be different from the normal orthorhombic foetal crystals, the habit being rather different and the direction of maximum pleochroic absorption (and of the higher refractive index) now being perpendicular to the needle axis instead of along it. Unfortunately, the crystals were far too small for X-ray examination, and after allowing them to stand for some days in the mother liquor it was observed that foetal crystals of the normal type (with maximum absorption along the needle axis) began to appear, and in the end completely replaced the first metastable form.

(*d*) *Foetus aged* 144 *days* (*sheep no.* 697)

The precipitate first obtained was apparently amorphous, and only after standing for 10 days did a few crystalline particles make their appearance; they were needle-like in form, but too small even to enable their optical characteristics to be determined.

Thus there are definite signs that a mixture of haemoglobins is present in foetuses aged more than 120 days.

Discussion

If two apparently different crystals have identical space-groups, cell dimensions and diffraction patterns, it is certain that they are composed of the same molecules; but if two crystals give different diffraction patterns all that can be said is that the molecules of which they are composed *may* be different. In the present case, therefore, it has been necessary to invoke other and less direct evidence. We have the circumstance that haemoglobins crystallized from a number of adult sheep give homogeneous crops of crystals which are identical both morphologically and as revealed by X-ray examination; while from different foetuses (of ages less than 120 days) we also obtained homogeneous crops of identical crystals, and these too maintained their identity on recrystallization. But the foetal crystal is always different from the adult type (in internal structure as well as in habit), and since the crystallizations of the two varieties were always carried out under exactly similar conditions, it seems reasonable to conclude that we are really dealing with two different proteins. The only doubtful point would appear to be the possibility of mixed crystal formation; but it is difficult to believe that mixed crystals could pass undetected after recrystallization, in view of the striking changes in behaviour observed when artificial mixtures were prepared. In any case even this hypothesis, which seems a most unlikely one, demands the presence of at least two proteins, in different proportions in the two cases.

Our findings with blood samples from late foetuses also conform to the general picture. We suppose that the foetal type of haemoglobin which alone is present in the blood of foetuses aged less than 120 days gradually gives place after this date to the adult type; the blood of late foetuses should contain a mixture of haemoglobins, and this is made very plausible by the similarity in behaviour on salting out between haemoglobins from late foetuses on the one hand, and artificial mixtures of adult and of early foetal specimens on the other. However, the timetable of the change-over remains unknown in detail.

There is, then, strong circumstantial evidence that foetal and adult haemoglobins, in sheep at any rate, are distinct chemical individuals. It remains to find the nature of the structural differences between them. There seem to exist certain differences in molecular symmetry which will be discussed in the next section. Apart from these there are no structural features which could be unravelled without pursuing the X-ray analysis to much greater detail than was practicable in this investigation. The X-ray work did, however, lead to certain results which are relevant to the general problem of haemoglobin structure, and these will now be considered.

5. FINDINGS RELEVANT TO GENERAL PROBLEMS OF HAEMOGLOBIN STRUCTURE

In this section are discussed a number of points which emerged in the course of the X-ray examination of adult and foetal sheep haemoglobins and which, while not all bearing directly on the primary problem of this research, do throw some further light on structural questions.

The densities of foetal and adult haemoglobin crystals: calculations of 'bound water'

One of us has discussed fully in two papers (Perutz 1946; Boyes-Watson *et al.* 1947) the significance of the swelling and shrinkage properties of haemoglobin crystals, the distinction between 'bound' and 'free' water, and the methods which may be used to evaluate the amount of the former. These methods were developed from those of Adair & Adair (1936) and of McMeekin & Warner (1942), modified to make use of the information provided by X-ray analysis. It has been shown that the amount of 'bound water' may be derived from a knowledge of the density of the wet crystal, of the partial specific volume of the anhydrous protein in solution, and of the dimensions of the unit cell. The density of the wet crystal may be determined by the method of McMeekin & Warner, while the partial specific volume has been shown by Adair & Adair (1936) to be nearly constant for a number of proteins, and to be the same no matter whether the protein is in solution or in the form of a crystalline suspension. The value for adult sheep haemoglobin is 0·7524 at 20° C.

In the present case it was possible to determine the densities of wet crystals of adult sheep haemoglobin and of monoclinic foetal haemoglobin; but for the orthorhombic variety of foetal haemoglobin consistent results were impossible to obtain by the method of McMeekin & Warner owing to the excessive thinness of the crystals. The values obtained in the former two cases, and the calculated value of w, the weight of 'bound water' per gram of protein, are shown in table 3. For comparison the table also gives the corresponding quantities for horse haemoglobin and human haemoglobin.

It will be seen that in spite of a fairly wide variation in total percentage of liquid of crystallization, the values of w are the same within the limits of experimental error. This is one of the few points of close resemblance so far discovered between the crystal structures of different haemoglobins, and it perhaps gains in significance from the fact that in the case of horse haemoglobin, where the overall molecular dimensions are known with some degree of certainty, a calculation of w from these dimensions on the assumption that the bound water forms a monomolecular layer over the whole surface of the protein molecule, gives a value of 0·28 g./g. protein. Within the limits of error this is identical with the experimentally determined values for all four proteins.

TABLE 3. DENSITIES OF WET HAEMOGLOBIN CRYSTALS AND WEIGHT OF BOUND WATER PER GRAM OF PROTEIN

type of crystal	medium	vol. % liq. in crystals	w (g./g. protein)
adult sheep	2·9 M-$(NH_4)_2SO_4$	55	0·29 ± 0·03
monoclinic foetal sheep	2·6 M-$(NH_4)_2SO_4$ + 0·02 M-$(NH_4)_2HPO_4$	41	0·27 ± 0·03
horse	2·5 M-$(NH_4)_2SO_4$ + 0·4 M-$(NH_4)_2HPO_4$	52	0·30 ± 0·03
human	4 M-$(NH_4)_2SO_4$ (neutralized)	41	0·26 ± 0·03

Evidence for intramolecular layering in foetal haemoglobin

The most striking feature of the diffraction pattern obtained from monoclinic foetal crystals is to be found in the $0k0$ reflexions. Figure 5, plate 10, shows a 5° oscillation photograph taken with the crystal rotating about c^* and the beam parallel to a; the central region contains the $0kl$ reflexions, and the equatorial spots in this region are the $0k0$'s. It will be observed not only that the odd orders are absent as demanded by the space-group, but also that 080 is almost absent and 060 completely so, while 040 and 0, 10, 0 are extremely powerful. Other pictures show that 020 and 0, 12, 0 are weak. The strikingly powerful 040 and 0, 10, 0 reflexions, and the relative weakness of other even orders, can only be due to a very marked layer structure running through the unit cell, the layers being normal to b and the projected electron density along this axis being essentially made up by Fourier summation of two components respectively one-fourth and one-tenth of the b axial length—i.e. 21 and 8·3 A. The contributions of the other visible $0k0$ reflexions would be very small, and a knowledge of the relative phases of 040 and 0, 10, 0 would make it possible to construct a fairly realistic one-dimensional projection of the electron density in the cell on to a line parallel to b; it is unfortunate that in this space-group the phases of the $0k0$'s are not necessarily 0 or π, so that there is no possibility of applying the same methods of analysis as were used by one of us (M.F.P.) in the case of the real $00l$'s of horse haemoglobin.

Since, however, there are only two molecules in the unit cell it seems impossible that strongly marked periodicities of 21 and 8·3 A could be produced *inter*molecularly; it seems reasonable to deduce that they are the expression of a well-developed *intra*molecular layering. Furthermore, since the layering runs right through the cell, the layers in the two molecules must be parallel to one another.

It has been pointed out that the monoclinic type of crystal exhibits very slight pleochroism and high absorption in all directions—in other words, the haem groups in the two molecules of the cell cannot be parallel. That is to say, the haem groups must be disposed at an angle to the layering, for if they were parallel to the layers strong pleochroism also should be exhibited (since the layers *are* parallel in the two molecules). The properties of orthorhombic foetal crystals agree with this scheme, since it is apparent from their strongly pleochroic properties that in them the haem groups are parallel throughout the cell, which means that the intramolecular layers cannot be so; and indeed the diffraction pattern exhibits no special absences among the lower orders which might be anticipated if layering throughout the cell existed.

It is interesting to see how the intensities of the $0k0$ reflexions change on drying. Figure 6, plate 10, is the dry crystal diffraction pattern corresponding with the wet picture of figure 5; it will be seen that 020 is now very powerful, 080 strong, and 040 fairly strong, while 060 is very weak. If the two original photographs are superimposed it can be seen that the wet 0, 10, 0 very nearly lies on top of the dry 080. This observation suggests that the layering postulated for the wet crystal persists in the dry state, and lends colour to the belief that—as in the case of horse haemoglobin—the layering is intramolecular and little affected by drying, though of course these points could not be proved without a knowledge of the phases and the construction of one-dimensional Fourier projections for each state.

The pronounced periodicity of 8·3 A should be compared with the spacing of just under 9 A found in the case of the four layers in the molecule of horse haemoglobin (Boyes-Watson *et al.* 1947).

Shrinkage experiments with monoclinic foetal haemoglobin

In the case of horse haemoglobin the study of the drying of the crystals was extended by isolating intermediate shrinkage stages, the technique being to open the capillary containing a mounted crystal for a certain time, and then to close it again and take an X-ray photograph; if the time was suitably chosen enough water had escaped for the crystal to shrink partially. Other shrinkage stages were obtained by opening the tube for further periods, until eventually the typical air-dried pictures were obtained. In this way two intermediate shrinkage stages were isolated.

This method has the limitation that the shrinkage is difficult to control, being secured merely by allowing an indefinite quantity of water to escape from the system. Furthermore, there is no means of determining at what vapour pressure the water escapes—for if, indeed, the water exists in the crystal in discrete layers, as the evidence indicates, each layer would be expected to come out at a characteristic vapour pressure.

A new method, which will be fully described elsewhere, was therefore devised in order to allow the shrinkage process to be carried out under controlled conditions. Its essence is to keep the crystal in an atmosphere of constant and known aqueous vapour pressure until equilibrium has been attained, and then to measure its cell dimensions and so forth. For this purpose the crystal is mounted in a thin-walled quartz tube as usual, and the latter is then inserted into the mouth of a glass chamber (about 1 cm. cut from the bottom of a small test-tube is suitable) and fixed there with plasticine, the whole being made sufficiently gastight by covering the plasticine plug with vacuum grease. Into the glass chamber are placed a few drops of a solution of known vapour pressure, and the whole device is then mounted on the goniometer head of the X-ray camera in the normal way. The solution found to be most suitable for the purpose is aqueous calcium chloride, since the vapour pressure of the saturated solution (about 6 mm. Hg at 20° C) is low enough to dry the crystal completely, so that by altering the calcium chloride concentration the entire needed range of vapour pressure can be covered. Furthermore, the crystal can be preserved over calcium chloride solutions for considerable periods without degeneration taking place.

In the case of monoclinic foetal crystals two calcium chloride concentrations were chosen, giving vapour pressures intermediate between those required to maintain the crystal in the fully wet state and to dry it completely. Table 4 shows the cell dimensions measured from crystals kept above these solutions; the fully dry and fully wet dimensions are also given for comparison.

TABLE 4. SHRINKAGE STAGES OF MONOCLINIC FOETAL HAEMOGLOBIN CRYSTALS

% $CaCl_2$ (by weight)	approx. vapour pressure (mm.)	*a* (A)	*b* (A)	*c* (A)	β	cell volume (A^3)
(fully wet crystal)		55	83	64	102°	286,000
23	13·5	49	77	59	102°	218,000
30	11·5	48	76	58	102°	207,000
(fully dry crystal)		48	71	58	101°	194,000

Although it is not possible in a case such as this (where high-order reflexions are no longer visible) to measure cell dimensions to the nearest Angstrom, owing to absorption effects and other errors, the small difference of 1 A between some of the figures in the table are real, since the photographs when superimposed show immediately the difference between the axial lengths as the spots drift out of perfect alinement across the picture. The intensities of the reflexions alter on shrinkage, as well as the cell dimenions; in other words there are at least four quite distinct lattices. It will be seen that as drying proceeds the three cell axes do not shrink simultaneously or to the same extent; *b* has shrunk by only just over half the full amount at the stage when *a* and *c* have shrunk to completion. This is more clearly shown in figure 7, where axial lengths and cell volume are plotted against vapour pressure. Unfortunately, the supply of monoclinic crystals was too small to enable us to make a study of their shrinkage in more detail, or, in particular, to discover whether, as seems to be the case for horse haemoglobin, the dimensions decrease in a series of jumps as the cell shrinks; on this hypothesis the curves of figure 7 would show a series of steps if analyzed more closely. It may, however, be observed that the decreases in cell dimensions fall into two groups—those of 1 A and those of 5 to 6 A. The latter correspond in magnitude to the stepwise changes in this thickness of the water layer of horse haemoglobin on shrinkage.

The diffraction pattern of orthorhombic foetal haemoglobin

The general features of the X-ray pattern of these crystals have been described in §3 (p. 382) above, and it has been shown that the cell contains four molecules of molecular weight 68,000, lying in special positions on diad axes. The photographs show further that, at least over most of the central regions, reflexions $0kl$ are absent when k is odd, whereas according to the space-group only those with l odd should be absent. Close examination shows that weak reflexions with k odd appear again further out in the diffraction pattern (e.g. 054, 074), and in one case among the low orders (012). The special absences among the low-order $0kl$'s lead to certain interesting conclusions which will now be considered.

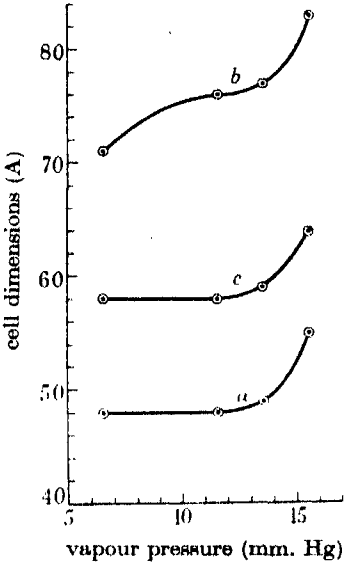


FIGURE 7. Shrinkage of cell dimensions of monoclinic foetal sheep methaemoglobin, as a function of water vapour pressure.

Normally, absences of the type $0kl$ with k odd would point to the presence of a glide plane. This solution would involve the turning of right-handed into left-handed molecules and can therefore be excluded in a protein crystal. It might be argued, however, that each asymmetric unit in the protein molecule may be treated as a mass of scattering matter within which the detailed stereochemical configuration does not matter so far as the low-order reflexions are concerned, and that these 'diffuse' masses of scattering matter may be related by 'pseudo'-glide planes. The presence of such planes would give rise to a pseudo-doubling of the number of asymmetric units in the cell. These units could be identical in their coarse features but differ in stereochemical detail.* This is the explanation adopted by Crowfoot & Schmidt (1945) for similar features in the diffraction pattern of a crystal of tobacco necrosis virus.

---

* There is the alternative possibility of placing molecules in special positions at $y = 0$ and $\frac{1}{2}$, but it can be shown that this does not lead to any useful results in the case of protein crystals.

In the present case an alternative procedure has been adopted, first because the concept of a pseudo-glide plane was thought to be unnecessarily complex, and secondly because it turns out that absences of the type $0kl$ with k odd could well extend over the entire range of $\sin\theta/\lambda$ without necessitating the introduction of any symmetry elements requiring right-handed molecules to be turned into left-handed ones. Our argument runs as follows.

In order to simplify our examination of the implications of the special absences we may suppose, in the first case, that they extend over the entire range of $\sin\theta/\lambda$ covered by the diffraction pattern. We may represent an asymmetric protein unit (of unknown size at this stage) by a (purely formal) asymmetric tetrahedron and place one such tetrahedron in a general position x, y, z and in general orientation in the unit cell, which is shown in projection along the a axis in figure 8 a and b. If we now allow the symmetry elements of the space group to operate on the tetrahedron at x, y, z, we obtain the set of eight tetrahedra shown in bold lines in figure 8b. These eight tetrahedra, of course, give rise only to the space-group absences and not to the special absences of $0kl$ with k odd. In order to account for the latter we must add a second set of eight tetrahedra (shown in faint lines in figure 8b), such that for each tetrahedron at x, y, z there is another with z unchanged, y changed to $y+\frac{1}{2}$ and x changed by an arbitrary amount $x+x'$. In effect another lattice is introduced which is related to the first by translations of $\frac{1}{2}y$ and $x+x'$.

The additional set of eight tetrahedra has not introduced any new symmetry element into the pattern or altered the size of the unit cell, yet it shows clearly that a minimum of sixteen identical protein units is required in order to produce the special absences among the $0kl$ reflexions. Since the number of molecules per unit cell is only four, this means that each of them must be composed of four identical subunits. The fact that some reflexions with k odd are actually present in the higher orders may mean either that the two sets of eight units differ in fine structure or that their relative positions and orientation are nearly but not exactly determined by the translations mentioned above.

Summing up, we may conclude that each molecule of weight 68,000 consists of two exactly identical units of 34,000, and that each of these in turn is composed of two units of molecular weight 17,000. The latter may either be similar in their general structural features, but differ in their finer stereochemical details, or else they may be identical in all respects, but occupy positions slightly different from those dictated by the absent reflexions. In any case the molecule must consist of four subunits of at least approximately equal molecular weight of 17,000. The four subunits may lie at the corners of a square or rectangle (if $x = -x'$) or of a regular tetrahedron or at some intermediate position. Any such arrangement would be compatible with the symmetry elements, and so far the X-ray data offer no clue as to the exact arrangement of the subunits.

The molecular symmetry of haemoglobins

It has been known for some time that, like certain other proteins, a number of haemoglobins split into smaller subunits in dilute solution or on addition of urea. Thus Wu & Yang (1932) have shown by the osmotic pressure method that in urea solution the haemoglobins of horse and ox dissociate into fragments of molecular weight about 34,000, while those of dog and (adult) sheep do not. In the case of horse this result was confirmed by Steinhardt (1938), who also showed the splitting to be reversible, re-association taking place on removal of the urea by dialysis. Simple dilution may produce the same effect, as was shown by Pedersen (1933)

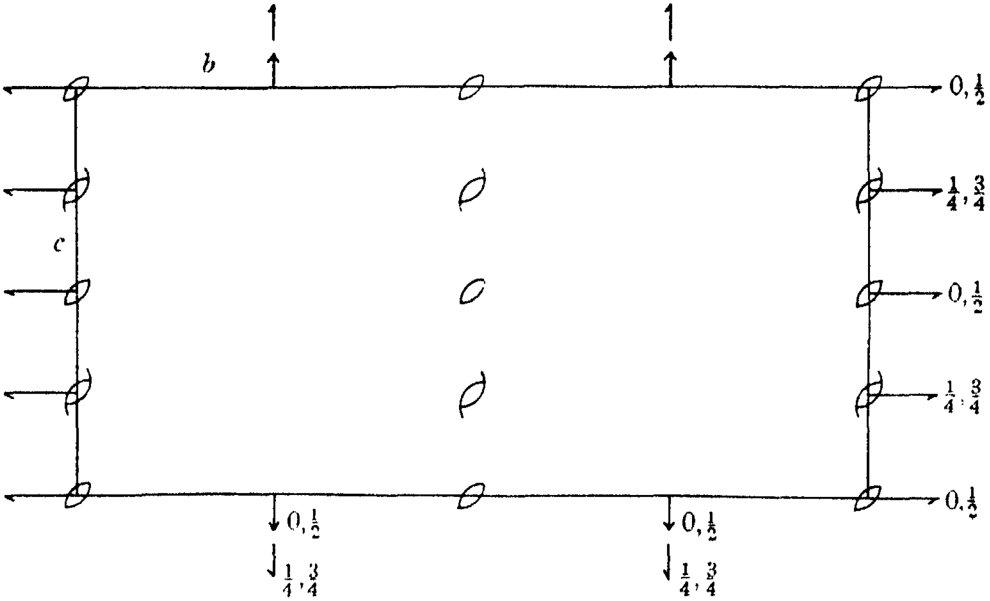


FIGURE 8*a*. Symmetry elements in the space-group $B22_12$ projected on to the *a* plane. The symbols and conventions conform to those of the *International Tables* (1935).

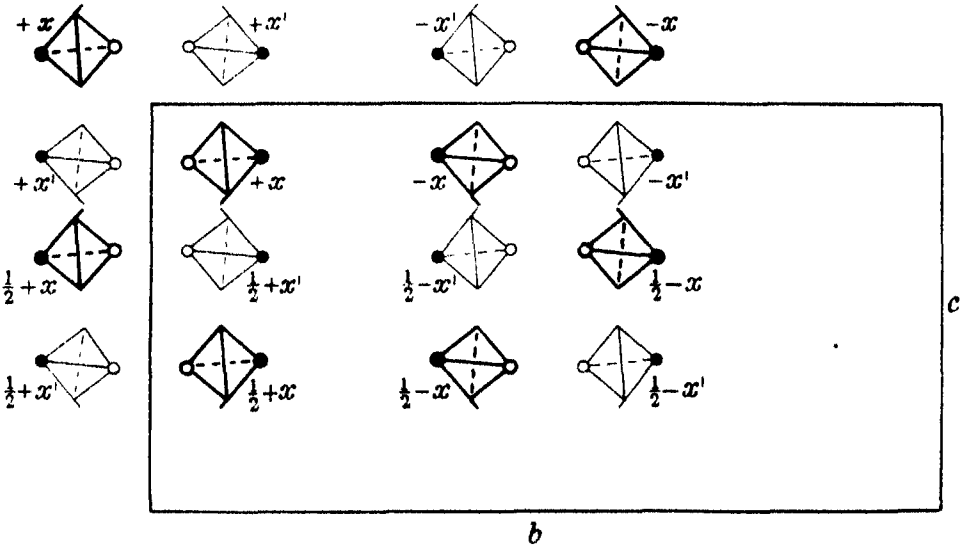


FIGURE 8*b*. Arrangement of asymmetric units in projection on the *a* plane. For the sake of clarity units at equivalent positions around three of the corners of the unit cell have been omitted.

and by Tiselius & Gross (1934), who measured the sedimentation constant and diffusion constant respectively of horse haemoglobin at various dilutions. In both cases they found evidence of dissociation into subunits at concentrations less than about 0·5 %. Svedberg & Hedenius (1934) report an even more pronounced effect for cat haemoglobin.

More recently, Gutfreund (1946) measured the sedimentation constant as a function of concentration for horse haemoglobin and foetal sheep haemoglobin and found definite evidence of splitting in each case and particularly in the latter where it even seems likely that the dissociation may proceed beyond the stage of half-molecules of molecular weight 34,000 (there were, in fact, indications that in the most dilute solutions dissociation into quarter-molecules takes place); in the case of adult sheep haemoglobin, however (Gutfreund 1947), there was no evidence of splitting, in agreement with the results of Wu and Yang. Using another technique (osmotic pressure), McCarthy & Popják (1947) report similar results, that maternal sheep haemoglobin shows no tendency to dissociate at high dilution, but that haemoglobin from sheep foetuses of ages ranging from 80 to 112 days dissociates appreciably at concentrations below 2 %.

Turning now to the results of X-ray analysis we wish to suggest a correlation between the highest symmetry displayed by the molecules in the crystal lattice, and their powers of dissociation in solution; in other words, that molecules which are situated on axes of symmetry in the crystal lattice and which are thus shown to consist of at least two structurally identical parts, may dissociate in solution; furthermore, the size of the dissociation products seems to be the same as the size of the smallest asymmetric unit in the crystal. This statement implies that these subunits cannot be joined together by peptide or other covalent links (e.g. disulphide or ester linkages), but must be joined by bonds which are capable of dissociation in the presence of water. So far no protein has been shown to dissociate into subunits smaller than the smallest asymmetric unit in the crystal; this might, however, be a result of the very limited amount of work which has so far been done on this aspect of protein chemistry.

It does not follow that the molecule always displays its full symmetry in the crystal. Thus in both horse and foetal sheep haemoglobins there are two forms, in one of which the molecule lies on a diad axis, while in the other it is in a general position. For the present discussion it is the highest symmetry exhibited in any form which is relevant.

The following three cases exemplify this suggested correlation:

(1) *Horse methaemoglobin*

This protein crystallized in the monoclinic space-group $C2$, which has four general positions, with only two molecules per cell; the molecule therefore exhibits diad symmetry. Dissociation occurs both on dilution and in the presence of urea. (The orthorhombic form has space-group $P2_12_12_1$ with four molecules in general positions, no special symmetry of the molecules being exhibited.)

(2) *Foetal sheep haemoglobin*

The orthorhombic form of crystal, with space-group $B22_12$, has four molecules in a cell containing eight general positions; the molecule therefore exhibits diad symmetry. Pronounced dissociation occurs on dilution. (The monoclinic form, with space-group $P2_1$, has two molecules per cell in general positions, and reveals no molecular symmetry.)

In addition, we have shown that in the crystal the molecule must consist of four similar or identical subunits of weight 17,000 each. At high dilutions the ultracentrifuge results suggest that splitting into four subunits may take place.

(3) *Adult sheep haemoglobin*

The only crystalline form of this protein which has been discovered has the space-group $C2$, with four molecules in a cell containing four general positions. The molecules therefore exhibit no intrinsic symmetry, and pending discovery of other modifications, may provisionally be held to possess none. There is no evidence of dissociation either on dilution or in presence of urea.

The above results are suggestive, but clearly further evidence from other species and other proteins is required before the correlation can be regarded as established. In the latter connexion it is noteworthy that the insulin molecule, which displays trigonal symmetry in the crystal, dissociates into three equal parts in dilute solution at appropriate pH (Gutfreund 1948).

6. CONCLUSIONS

1. Adult sheep methaemoglobin forms monoclinic plates belonging to the space-group $C2$. Under identical conditions of crystallization methaemoglobin prepared from foetuses aged less than 120 days is precipitated either as orthorhombic needles (space-group $B22_12$) or as monoclinic plates belonging to the space-group $P2_1$. There is evidence that the two latter types of crystals are modifications of the same protein, and that this protein is different from the haemoglobin of the adult sheep.

2. In artificial mixtures of foetal and adult haemoglobin crystallization is largely inhibited. Those crystals which do appear are either of the adult or of the orthorhombic foetal type, depending on which protein forms the major part of the mixture. A similar inhibition of crystallization is shown by haemoglobin of foetuses approaching term. Thus there is evidence of a mixture of proteins being present in late foetuses, but the details of the change-over from foetal to adult haemoglobin are still obscure.

3. No general picture of the structural differences between the adult and the foetal haemoglobin has emerged from this study, mainly because such differences were obscured by differences in molecular packing. There were, however, certain indications of differences in molecular symmetry which could be correlated with differences in the dissociation properties of the two proteins in solution. Adult haemoglobin forms an asymmetric unit of molecular weight 68,000 in the crystal and shows no sign of dissociating in dilute solution; foetal haemoglobin, on the other hand, is shown to consist of four identical or nearly identical subunits of molecular weight 17,000 which are symmetrically related in the crystal, and correspondingly it tends to split into quarter molecules in dilute solution.

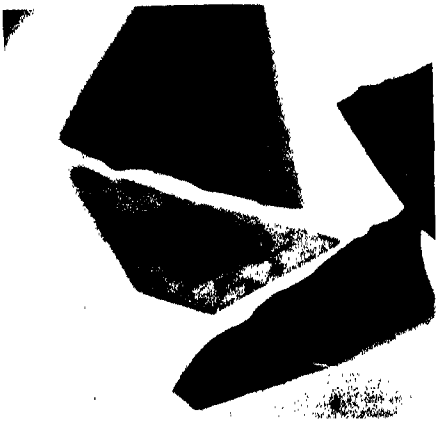


FIGURE 1

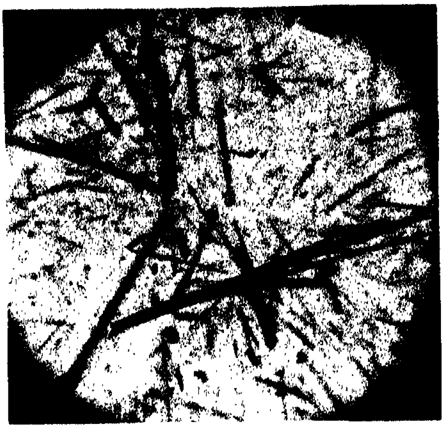


FIGURE 3

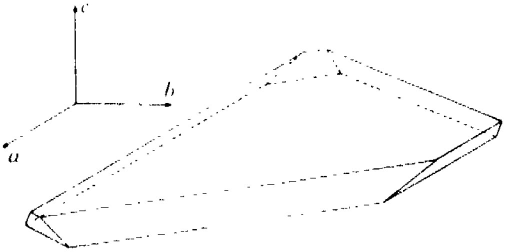


FIGURE 2

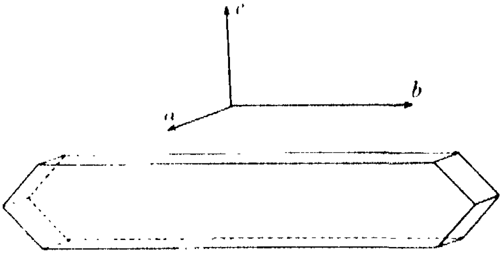


FIGURE 4

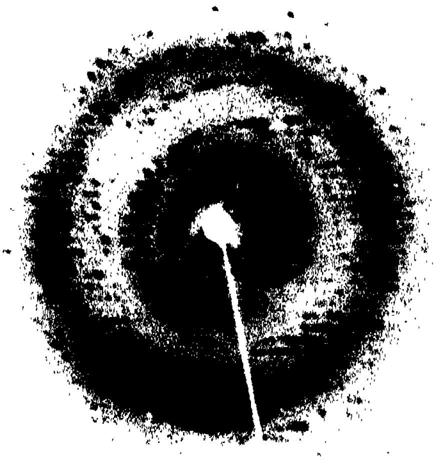


FIGURE 5

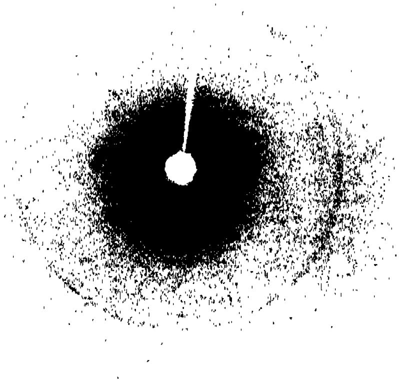


FIGURE 6

(*Facing p.* 397)

Monoclinic foetal haemoglobin shows evidence of an intramolecular layering with a period which is made up of the sum of two Fourier components, one of 21 and the other of 8·3 A spacing. The phase relation between these has still to be found. The four haem groups are probably parallel or nearly parallel in both the adult and the foetal haemoglobin molecule.

4. A new method of controlled shrinkage at constant humidity has been developed and applied to monoclinic foetal haemoglobin. Shrinkage was shown to proceed by transition between different sets of unit-cell dimensions, each of them stable at a definite humidity.

5. Both the adult and the foetal forms of sheep haemoglobin contain 'bound water' amounting to about 0·3 g./g. protein, which is of the same order as in haemoglobin of horse and man. The amount of 'bound water' does not depend on the total volume fraction of liquid present in the wet crystals.

The authors wish to make grateful acknowledgements to Sir Lawrence Bragg, F.R.S. and to Professor D. Keilin, F.R.S., for their advice and encouragement throughout the work; and in particular they wish to record their great debt to the late Sir Joseph Barcroft, F.R.S., at whose suggestion this research was carried out and who maintained active interest in its progress up to the time of his death.

DESCRIPTION OF PLATE 10

FIGURE 1. Crystals of adult sheep methaemoglobin. × 30 approx.

FIGURE 2. Clinographic drawing of adult sheep methaemoglobin crystal.

FIGURE 3. Crystals of orthorhombic foetal sheep methaemoglobin. × 30 approx.

FIGURE 4. Clinographic drawing of orthorhombic foetal sheep methaemoglobin.

FIGURE 5. X-ray diffraction picture of wet monoclinic foetal sheep methaemoglobin showing $0kl$ reflexions near the centre. 5° oscillation about c^*, beam parallel to a, specimen-plate distance 6 cm., Cu $K\alpha$ radiation.

FIGURE 6. X-ray diffraction picture of dry monoclinic foetal sheep methaemoglobin. Details as in figure 5.

REFERENCES

Adair, G. S. 1928 *Proc. Roy. Soc.* A, **120**, 573.
Adair, G. S. & Adair, M. E. 1936 *Proc. Roy. Soc.* B, **120**, 422.
Adair, G. S. 1939 Unpublished.
Andersch, M. A., Wilson, D. A. & Menten, M. L. 1944 *J. Biol. Chem.* **153**, 301.
Barcroft, J. 1935 Croonian Lecture *Proc. Roy. Soc.* B, **118**, 242.
Barcroft, J. 1946 *Researches on pre-natal life.* Oxford: Blackwell.
Boyes-Watson, J., Davidson, E. & Perutz, M. F. 1947 *Proc. Roy. Soc.* A, **191**, 83.
Brinkman, R., Wildschut, A. & Wittermans, A. 1934 *J. Physiol.* **80**, 377.
Brinkman, R. & Jonxis, J. H. P. 1935 *J. Physiol.* **85**, 117.
Brinkman, R. & Jonxis, J. H. P. 1936 *J. Physiol.* **88**, 162.
Crowfoot, D. 1938 *Proc. Roy. Soc.* A, **164**, 580.
Crowfoot, D. & Schmidt, G. M. J. 1945 *Nature*, **155**, 504.
Darrow, R. R., Nowakowsky, S. & Austin, M. N. 1940 *Arch. Path.* **30**, 875.
Gutfreund, H. 1946 Personal communication.
Gutfreund, H. 1947 Personal communication.
Gutfreund, H. 1948 *Biochem. J.* **42**, 544.

Haurowitz, F. 1930 *Hoppe-Seyl. Z.* **186**, 141.
Haurowitz, F. 1935 *Hoppe-Seyl. Z.* **232**, 125.
International tables for the determination of crystal structures. 1935 Berlin.
Karvonen, M. J. 1947 Personal communication.
Keilin, D. & Hartree, E. F. 1935 *Proc. Roy. Soc.* B, **117**, 1.
von Krüger, F. 1925 *Z. vergl. Physiol.* **2**, 254.
von Krüger, F. & Bischoff, H. 1925 *Ber. ges. Physiol.* **32**, 696.
von Krüger, F., Bischoff, H. & Schulte, H. 1926 *Ber. ges. Physiol.* **37**, 133.
McCarthy, E. F. & Popják, G. 1943 *Biochem. J.* **37**, xviii.
McCarthy, E. F. & Popják, G. 1947 *Nature*, **159**, 198.
McMeekin, T. L. & Warner, R. C. 1942 *J. Amer. Chem. Soc.* **64**, 2393.
Nicoletti, F. 1930 *Arch. antrop. crim.* **50**, 386.
Nicoletti, F. 1932 *Amer. J. Dis. Child.* **43**, 719.
Pedersen, K. O. 1933 Unpublished.
Perrier, C. & Jannelli, P. 1931 *Arch. Fisiol.* **29**, 289.
Perutz, M. F. 1946 *Trans. Faraday Soc.* **42** B, 187.
Perutz, M. F. & Weisz, O. 1947 *Nature*, **160**, 786.
Porter, R. R. & Sanger, F. 1948 *Biochem. J.* **42**, 287.
Reichert, E. T. & Brown, A. P. 1909 *The crystallography of the haemoglobins.* Washington Carnegie Trust.
Steinhardt, J. 1938 *J. Biol. Chem.* **123**, 543.
Svedberg, T. & Hedenius, A. 1934 *Biol. Bull. Woods Hole*, **66**, 191.
Tiselius, A. & Gross, D. 1934 *Kolloidzschr.* **66**, 11.
Wu, H. & Yang, E. 1932 *Chin. J. Physiol.* **6**, 51.
Wyman, J., Rafferty, J. A. & Ingalls, E. N. 1944 *J. Biol. Chem.* **153**, 275.

Laboratory experiments on whirlpools

BY A. M. BINNIE AND G. A. HOOKINGS
Engineering Laboratory, University of Cambridge

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 24 *January* 1948)

Two vertical pipes, placed on the axis of an open circular tank and fitted with trumpet entries of different shapes, were used in turn as overflows. Water was led into the tank in such a way that the stream approaching the trumpet under test possessed tangential velocity, which caused a vortex to appear in and above the trumpet. The relations between head and discharge were determined under various conditions of tangential supply. The formation of a vortex greatly reduced the discharge, the decrease being even more marked with the shallow than with the deep trumpet. At low heads, where the flow was controlled by the weir action of the trumpet crest, the discharge varied with the shape of the crest as well as with the irrotational constant c and the head, therefore (as with purely radial supply) it must be determined by experiment. At high heads with the trumpet flooded, the throat controlled the flow; within certain limits, the discharge was dependent chiefly upon c and the head measured with the throat as datum, and an approximate theory which ignores friction was verified. Two types of instability, 'surging' at low heads and 'spluttering' at high heads, were examined. Both were due to the collapse and subsequent re-formation of feeble vortices, which caused the discharge to vary with time in a periodic manner.

1. INTRODUCTION

The behaviour of whirlpools set up in the water at the submerged entrance to a pipe has long been regarded as obscure. Of recent years the problem has become important because of the increasing use of trumpet-shaped circular weirs for removing surplus water from reservoirs. Published information on the subject is scanty, and what little exists deals principally with the coefficients of discharge of trumpets having various shapes and sizes and with methods of suppressing the vortex at entry. G. M. Binnie (1938), who was concerned with the design of trumpets for the Jubilee Reservoir, Hong-Kong, used numerous models, most of which were very large judged by laboratory standards. But the arrangement of the lake constructed for the purpose, in which the models were placed, precluded control of the tangential velocity (or 'swirl') of the water approaching the trumpet, therefore no correlation was established between the swirl and the discharge through the trumpet. In the following pages an account is given of an attempt to elucidate this point and to determine the general characteristics of flow in the presence of vortices. The scale adopted was small, and it is not to be expected that the actual numerical results can serve as a highly accurate guide to the performance of a large installation.

The trumpet under test was placed in the centre of a circular open tank to which water could be supplied (i) radially, or (ii) through pipes set tangentially, or (iii) by means of a combination of (i) and (ii). Thus some measure of control over the swirl was possible, and the reduction in the discharge due to vortex formation could be determined. Experiments were made over a wide range of heads corresponding to the action of a reservoir trumpet under both normal and catastrophic conditions; the range was sufficiently great to cover also the case of a trumpet used as an overflow in a tank, its crest being placed considerably below the top of the tank. For purposes of comparison tests were carried out also on a pipe of uniform diameter. No attempt was made to suppress the vortex by any method such as the insertion of radial piers on the crest of the trumpet, thus the results may perhaps be regarded as purely cautionary—demonstrating the types of instability caused by feeble vortices, and also the enormous reduction in discharge due to vortex formation, which was briefly noted by Gibson (1911). However, it was thought that interest in the action of whirlpools was amply sufficient to justify the work. The experiments may be regarded as a natural extension of those described by A. M. Binnie (1938) and by A. M. Binnie & Wright (1941) who, using similar apparatus except for the tangential supply pipes, confined their attention to conditions of radial flow.

It should be added that no great progress has hitherto been made by purely theoretical methods. Elementary considerations suffice to show that in a perfect liquid of infinite extent the form of the surface of a free vortex is hyperbolic provided that the vertical velocity is zero everywhere (Lamb 1932). But if discharge occurs through a vortex situated above a trumpet, the question becomes one of the notoriously difficult class known as free-surface problems. The recent development of relaxation methods by Southwell (1946) provides a means of solving particular cases, and an account will be published elsewhere of the application of these methods to vortex flow through an orifice-plate; in that investigation the motion has been taken as *irrotational* so that (in accordance with the principle of the conservation of angular momentum) the swirl v at radius r was given by $v = c/r$, where c was the same for every particle of the liquid. But in the present experiments the effects of friction were found to be important, not only over the inner surface of the trumpet where the swirl was very great but also throughout the circular tank.

2. DESCRIPTION OF THE APPARATUS

The plan and half-sectional elevation of the apparatus are shown in figure 1. The circular tank *A* was supported on three levelling screws, and over a circular hole in its bottom the brass casting *B* was bolted. The plate *C* closed the base of the casting, and the stainless steel pipe *D*, to the top of which the trumpet was fitted, was soldered to this plate. For radial supply to the tank, the water was admitted through a control valve and a symmetrical arrangement of 2 in. piping to two openings diametrically opposed in the casting *B*. It then passed vertically upwards through the annular space in the casting *B*, in which radial fins were placed to destroy any rotation about the vertical axis. After emerging through the bottom of the tank, the water was forced to flow outwards by the baffle plate *E* which was supported on three small feet, thus the incoming water could not disturb the flow near the trumpet. To check this arrangement, drops of malachite green solution were inserted at various points in the tank, and it was seen that the supply to the trumpet was steady and truly radial. For the tangential supply the water passed through a control valve, an

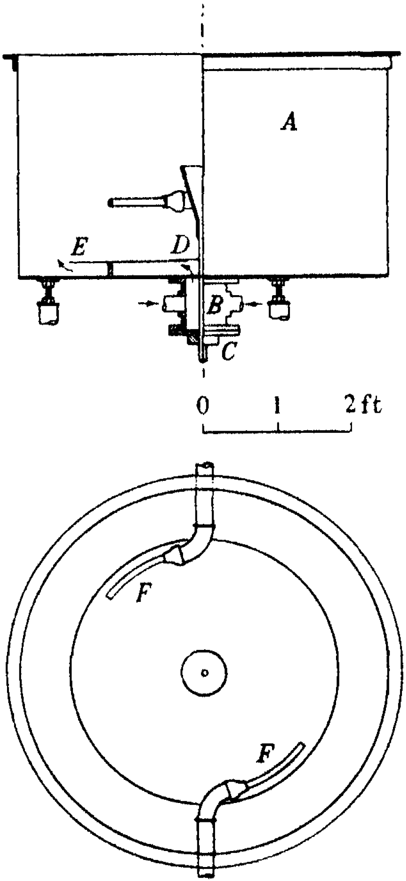


FIGURE 1. Plan and elevation of circular tank.

orifice-plate and symmetrical piping to the two pipes F which were bent to the correct radius. Three pairs of pipes F were used in turn, their nominal diameters being $\frac{1}{2}$, 1 and 2 in.

After leaving the circular tank through the pipe D the water (together with any entrained air) fell into the bell G (figure 2), which was clipped to the lower end of the pipe and which was supported in the open tank H. Here the air and water were separated, the water passing over the weir J to a calibrated measuring tank. The air was extracted from the bell through the steadying vessel K and the gas meter L by means of the fan M, which could be controlled with the air of the gauge N so that the pressure in the bell remained exactly atmospheric.

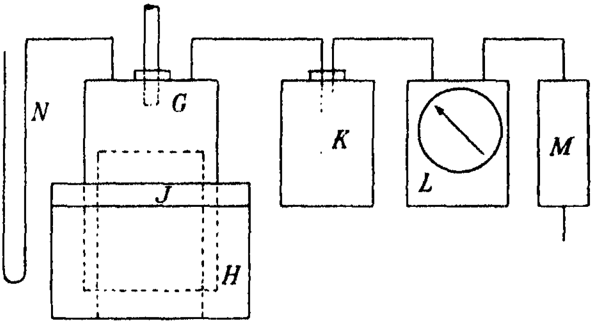


FIGURE 2. Air measurement apparatus.

A hook gauge with a range of 19 in. was used for measuring the water level in the circular tank, its zero being determined by means of a subsidiary hook gauge in the manner described by A. M. Binnie & Wright (1941). In these gauges the arrangements for supporting the hooks were of a special design, and an account of them may be of interest. The type commonly used in laboratories consists of some form of slider actuated by a screw or a rack. This system, although satisfactory when new, develops a serious defect after a little use. It is then found that the inevitable wear of the slider causes the normal vertical movement of the tip of the hook to be accompanied by a most undesirable lateral motion or shake. To overcome this difficulty the pattern shown in figure 3 was evolved. It is of correct geometric design so that no shake is possible, and it is easy to construct for no gear cutting is required. The casting a, which resembles a four-legged table on its side with part of the top cut away, is fixed and supports a vernier, and through its legs slides the graduated tube b carrying the hook. The double conical wheel c serves a treble purpose: (i) it is pressed against the tube by the spring washers d at its supports, thereby securing firm sliding contact at the points e, e', f and f' which remove four degrees of freedom; (ii) because it fits over the rail g fixed to the tube, it destroys the fifth degree of freedom of the tube, namely, rotation about the vertical axis; (iii) when rotated by means of the operating wheel h, it moves the tube vertically, which is the only direction in which motion is possible. The jamming action of the conical wheel on the rail ensures that the friction force there considerably exceeds the sum of the friction forces at e, e', f and f', and thus the tube can be raised by turning the operating wheel. To prevent the tube from descending when released, sufficient friction can be obtained at the sliding surfaces and at the bearings of the spindle on which the conical and operating wheels are mounted.

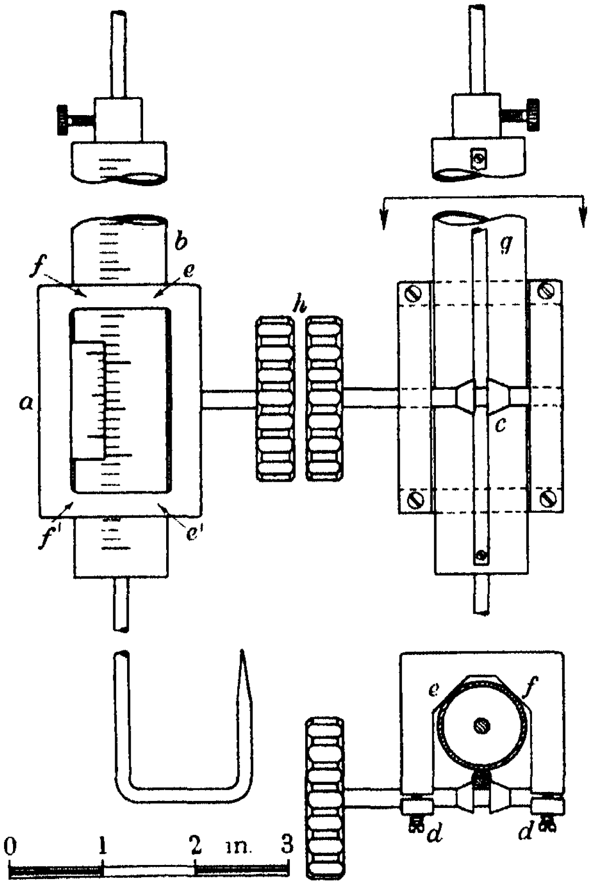


FIGURE 3. Hook gauge.

The trumpets used, which are shown in figure 4, were machined gun-metal castings push-fitted over the vertical pipe *D* indicated in figure 1. The plain pipe which was also tested had its upper end chamfered downwards and outwards at 45°. Each of these three arrangements had a total overall height of 33 in. from crest to outlet so that a fair comparison could be made between them. Their crests were fixed 20½ in. above the base of the tank in order to minimize the influence of the bottom. The trumpets were very nearly geometrically similar to those employed by G. M. Binnie (1938) and by A. M. Binnie & Wright (1941). In the deep bellmouth *A* the low-pressure region is far from the surface in the tank and vortices are not readily formed in it, but it is of a shape too costly to construct on a large scale. Bellmouth *B* is of practicable form for use in reservoirs, although its action is liable to be disturbed by a small swirl in the supply.

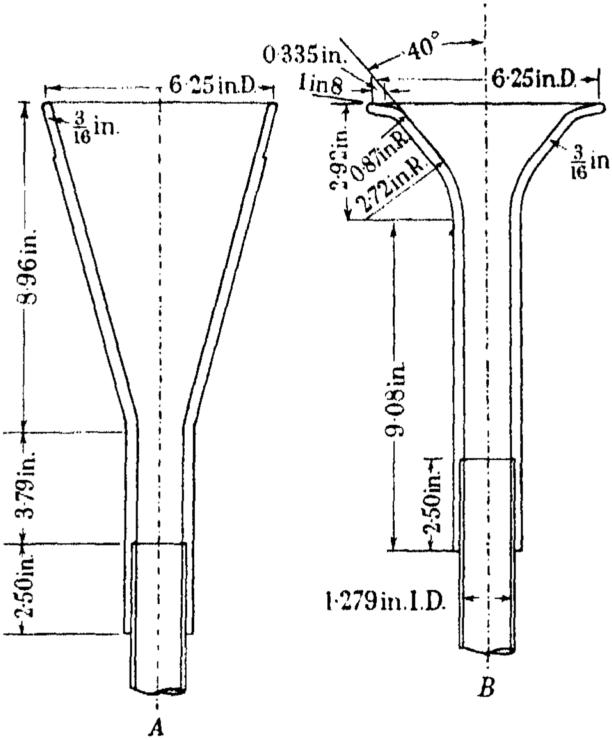


FIGURE 4. Bellmouths *A* and *B*.

3. DISCHARGE TESTS

As a first step the relation between the discharge Q and the head H above the crest was determined for each overflow, the radial and the three tangential supplies being employed in turn. The results are shown in figures 5 to 7. The typical dimension of each overflow is taken to be the diameter $2a$ of the throat and tail pipe, hence the non-dimensional quantities plotted are $H/(2a)$ and $Q/\{(2a)^{\frac{5}{2}}(2g)^{\frac{1}{2}}\}$. The earlier investigations with radial supply showed that under certain circumstances the Q-H relation was different with rising and falling values of H. This point was again examined, and where a difference was found it will be commented upon. A full description was previously given of the action of the three overflows with radial supply, and it need not be repeated here. It is sufficient to state that with the slightly larger overflows now employed the same behaviour was observed, and that the supply was so free from swirl that the few vortices seen were small and transient.

Plain pipe (figure 5)

With the tangential supplies, the discharge at very low heads differed little from that measured with the radial supply, and the flow was controlled by the weir action of the top of the pipe. An air core occupied most of the interior of the pipe, and the water flowed down helically in contact with the wall, emerging from the outlet as a spinning annular jet. When Q was increased, the critical conditions were soon reached where the control of the discharge shifted to the carrying capacity of the pipe, and here the experimental curves turned steeply upwards. The very large reduction in the critical values of Q and H compared with those obtained with radial supply was due to two causes: (*a*) since a permanent air-core occupied part of the interior of the pipe, the cross-section available for the water flow was decreased and the length of the pipe had no influence upon the discharge; (*b*) the velocity of the water in the pipe due to its loss of height had of necessity a large tangential component in the horizontal plane due to the small radius (in plan) of its path, hence its vertical component was greatly diminished. In all these tests the air-core of the permanent vortex reached from the surface in the circular tank down through the pipe. As it entered the pipe its diameter increased because in this region the vertical acceleration of the water was considerable and a corresponding reduction in the cross-section of the stream occurred. The vertical velocity of the water increased as it passed down the pipe, causing a further enlargement of the air-core diameter. Unless the latter was very small, its axis was almost perfectly steady, but the surface of the core was disturbed by ripples of helical appearance, which in a powerful vortex rose spinning with sufficient intensity to be visible even on the almost horizontal surface in the tank. There they often gave a casual observer a false impression of the direction of rotation of the vortex. The existence of these ripples may be explained by reference to the work of Rayleigh (1892), who showed that, owing to surface-tension effects, the surface of a cylindrical air-core in a liquid is unstable and subject to varicose undulations of wave-length about 6·48 times the core diameter. This result is true also when the liquid is spinning irrotationally. It appears that, when corrugations of this type are formed on the rotating and falling trumpet-shaped surface of an actual vortex, they give rise to the system of waves described above.

Tests were made with combinations of the radial and tangential supplies, yielding curves intermediate between those shown in figure 5. They were carried out by setting the tangential supply at a fixed value and then altering the radial supply in steps. This procedure led to the discovery of a type of instability due to the formation and collapse of very feeble vortices. It was found with the 2 in. tangential supply in use, which for an obvious reason produced a thinner vortex than the other two. As the radial supply was increased, thus reducing the swirl, the vortex became still more slender until eventually the stage was reached where a permanent air-core could no longer be maintained. This point was well defined, for a loud spluttering noise was emitted due to the rapid opening and shutting of the air-core. The level in the tank immediately commenced to fall, indicating that the greater discharge had begun which is associated with coreless flow. After a time the frequency of the spluttering decreased, and finally stopped when a permanent air-core was re-established. Thereafter the water level again rose, and as soon as it had regained its former value, the cycle, which had a period of about 45 min., was repeated. This succession of events is explained by figure 8, which shows the actual values of Q and H recorded in the relevant tests. For a constant tangential supply of 15·0 cu.in./sec. the curve consists of two disconnected parts: (i) DA, representing the

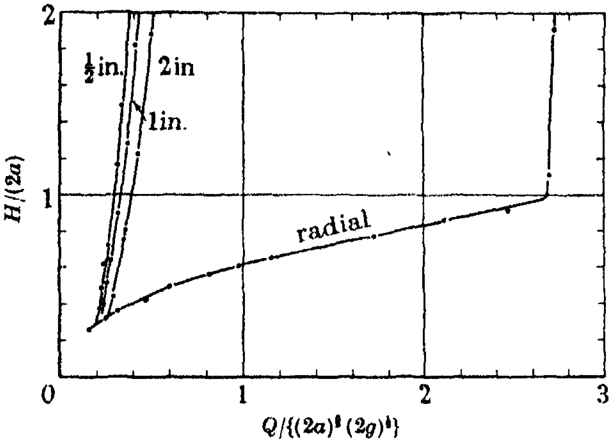


FIGURE 5. Discharge tests on the plain pipe.

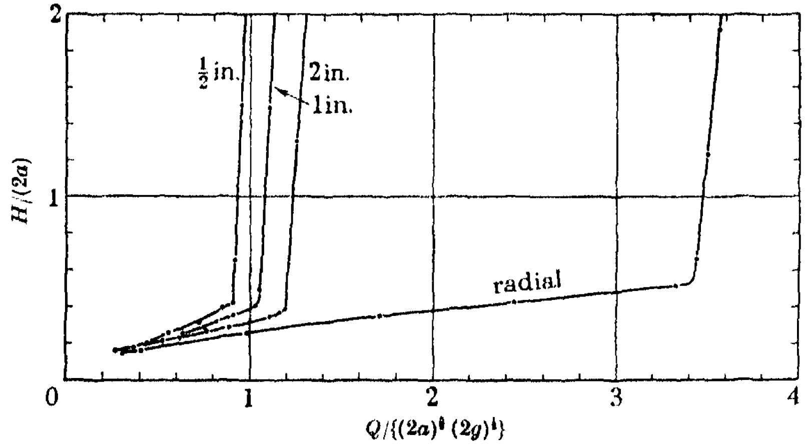


FIGURE 6. Discharge tests on bellmouth *A*.

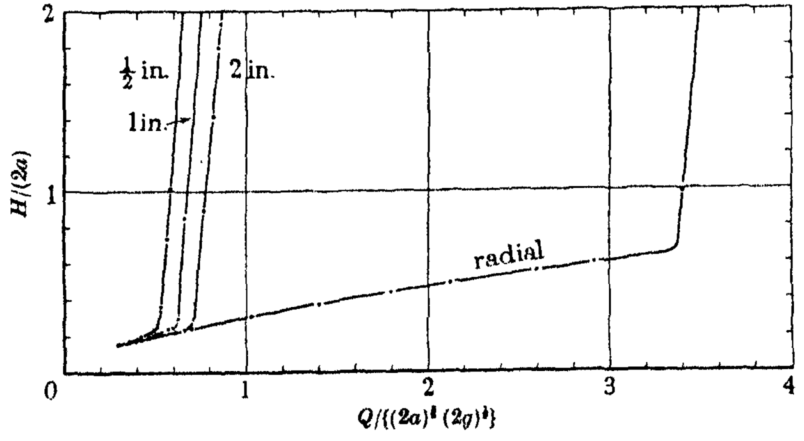


FIGURE 7. Discharge tests on bellmouth *B*.

flow with an air-core which was obtained by commencing the test at zero head; (ii) *BC*, for the coreless flow which resulted on beginning at a very high head. The curves could be traversed with both rising and falling heads provided that the upper limit *A* and the lower limit *C* were not passed. Accordingly, if the point *A* was attained in the sequence mentioned above, the discharge changed suddenly to that indicated by *B*. This was much larger than the supply, consequently the level fell to a point near *C*, which gave the lowest value of *H* at which coreless flow was possible. Thereupon the discharge moved to a point near *D* and was now smaller than the supply, therefore the level rose to *A*, thus completing the cycle. Figure 8 also shows that a similar but less marked effect occurred with the larger vortex associated with a tangential supply of 22·9 cu.in./sec.

Bellmouth A (figure 6)

With the tangential supplies it was again found that at very low heads the effect of the swirl upon the discharge was slight. This was confirmed by observations of surface particles and filaments of coloured liquid, which moved in almost circular paths until they were very near the crest. Once they came within the immediate influence of the crest, the direction of their paths changed sharply by nearly 90°;

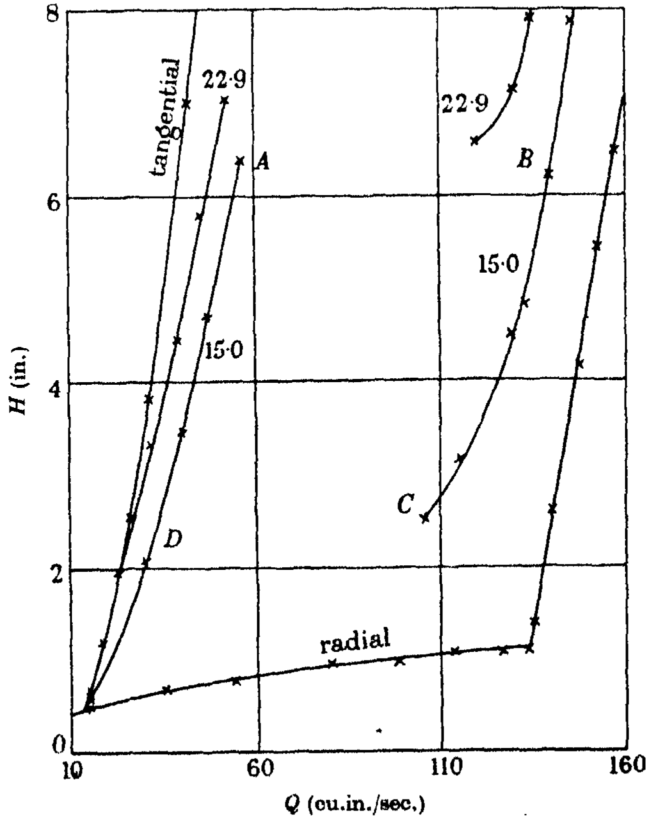


FIGURE 8. Discharge tests on the plain pipe showing instability due to spluttering.

and since they passed over the crest almost radially, the discharge was affected only slightly by the original swirl. As the head was increased, the onset of critical conditions was shown by the appearance in the trumpet of a shelf of rotating water with an air-core. A further rise in the head brought the shelf up to the surface in the tank, the control of the flow then passing from the trumpet crest to the throat.

The radial and tangential supplies, used together in the manner described above, again led to interesting results. With the ½ in. supply set at 15·6 cu.in./sec. the normal curve of weir action was obtained until the shelf appeared when the radial supply was 64·8 cu.in./sec. But under these conditions the level of the shelf, instead of remaining constant, rose and fell in the trumpet. At a radial supply of 81·4 cu.in./sec. this surging was more marked, the fluctuations extending from the crest of the trumpet almost to the throat. A further increase of only 2·4 cu.in./sec. produced a marked change, for the level in the tank suddenly rose by nearly 5 in. as indicated by the broken line *BC* in figure 9. Higher rates of flow produced the curve *CD*, but when the radial supply was decreased, the observations followed the course *DCEA* back to the weir portion of the curve, showing that the high-head vortex once formed was stable. These effects did not occur when the tangential supply was 21·9 cu.in./sec.

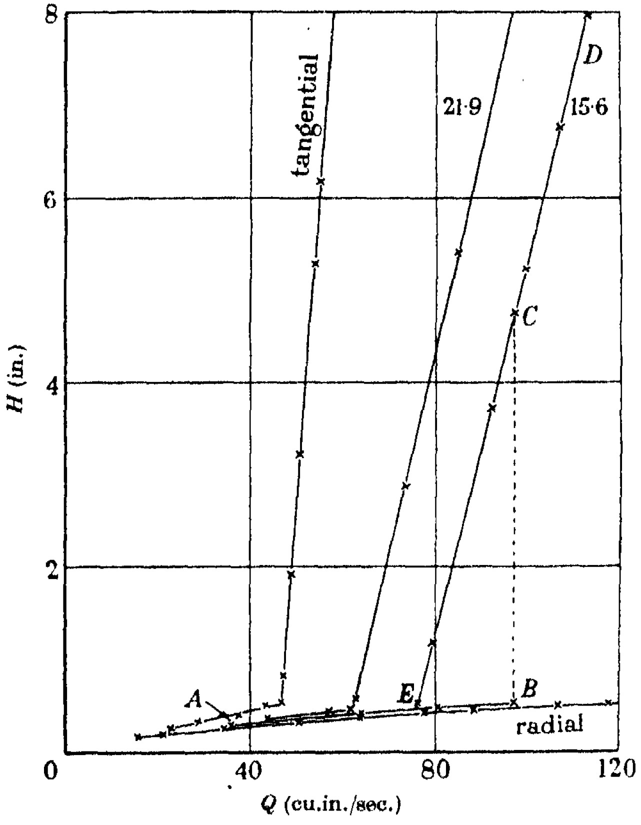


FIGURE 9. Discharge tests on bellmouth *A* showing instability due to surging.

or higher. With the 1 in. tangential supply the surging ceased before the trumpet was completely filled up, but the ensuing rise in level, consequent on the formation of a high-head vortex, was so great that it exceeded the 15½ in. freeboard between the trumpet crest and the top of the tank. The use of the 2 in. pipe had more intricate consequences. With the tangential supply set at 12·0 cu.in./sec. the vortex in the partially filled trumpet was so feeble that it spluttered, emitting a loud rasping noise and entraining considerable volumes of air; and when the radial supply was further increased the flow became entirely coreless. At 15·0 cu.in./sec. surging occurred, and the sudden rise in level which followed the filling up of the trumpet was about 7 in. But the resulting vortex was so feeble that spluttering ensued and the curve corresponding to *DCE* in figure 9 could not be followed. Instead, the level in the tank fell and rose in the manner described in connexion with figure 8. At 28·9 cu.in./sec. the surging was followed by so violent a rise in level that the radial supply had hastily to be reduced, but the vortex proved to be stable, and a curve similar to *CE* in figure 9 was recorded.

Useful evidence on surging action was obtained by employing a Perspex trumpet, similar in shape and size to bellmouth *A* and placed on top of a glass tube. It was seen that the fall in level in the trumpet was accompanied by the passage through the pipe of a coreless block of water. The core was immediately re-established, but as the discharge capacity of the pipe under these conditions was less than the supply over the trumpet crest, the level in the trumpet began to rise. But, due to friction at the wall and to shock where the falling water impinged on the shelf, the energy loss in the partially filled trumpet was sufficient to prevent the maintenance of the core. Consequently the discharge capacity of the pipe suddenly increased, and the contents of the trumpet were again gulped down the pipe.

Bellmouth B (figure 7)

A comparison of figures 6 and 7 shows that, with radial supply, critical conditions were reached at almost the same discharge although at different heads; at high heads the shape of the trumpet had little influence. The expectation was confirmed that in the presence of vortices the discharge of bellmouth *B* was much the lower of the two. The main point of interest in this set of experiments was, however, the spectacular nature of the surging which was observed within a small range of supply conditions. With the 2 in. supply set at 12·1 cu.in./sec. surging set in with a radial supply of 51·9 cu.in./sec. This quickly increased in violence until the filling up of the pipe after a gulp was so rapid that the water issued from the throat in the form of a rotating jet, which, owing to centrifugal action, spread out like a mushroom. It then fell back into the throat and at once disappeared. The surging ceased when the radial supply was increased above 60 cu.in./sec. Thus the evidence suggests that surging may be expected when the supply to the trumpet has a low and critical amount of swirl. If the swirl is too small, no core is formed in the partially flooded trumpet; if too great, the vortex is sufficiently powerful to persist when the trumpet is completely filled.

Surging with a similar but much larger trumpet was noticed by G. M. Binnie (1938), but not on so marked a scale. This was perhaps due to the fact that his model was fitted with an inclined tail pipe. Thus after a gulp the air-core could not be so readily re-established as when the vertical pipe below the trumpet terminated in an open end.

4. VARIATION OF c IN THE CIRCULAR TANK

For a viscous liquid rotating about a vertical axis it can be seen from the equations of motion (Goldstein 1938) that, if the vertical velocity is zero everywhere, a distribution of swirl given by vr = constant is possible whatever the radial velocity may be. Thus, by itself, viscosity does not cause changes in the magnitude of the constant. Provided that the inlet disturbances are not serious, the above relation may be expected to hold good at levels in the tank far from the bottom and at radii intermediate between the supply pipes and the trumpet. The investigation of this point proved difficult. The velocities were too low for a Pitot tube to be employed except near the air-core, and there its use upset the vortex. Drops of coloured liquid introduced into the tank were also unsatisfactory, for they were quickly drawn out into long spiral filaments. The method finally adopted was to time the rotation of a match stick loaded with lead at one end and floating almost totally immersed with its long axis vertical. When steady motion had been established, centrifugal action constrained the rod to rotate in an almost circular path, the radius of which could be changed by altering the total weight of the rod. The 2 in. supply was found satisfactory, for at 58·5 cu.in./sec., the largest flow which was examined, c was found to be 27 in.^{2}/sec. at radius 17¼ in. and 22 in.^{2}/sec. at radii 13, 7½ and 3¼ in. At smaller radii the motion of the rod was too precarious for the method to be practicable, but the effective value of c at the throat was deduced from pressure measurement, as explained in § 5, to be 23 in.^{2}/sec. On the other hand, with the ½ in. pipes, with which the inlet disturbances and the speed of rotation were greater, c fell from 88 in.^{2}/sec. at radius 18 in. to 44 in.^{2}/sec. at radius 7 in. when the flow was 66·4 cu.in./sec., showing that this method of supply was too violent to be entirely satisfactory. The experiments also indicated how great a loss of energy occurred in the tank. The ratio of the observed swirl at radius 18¼ in. to that calculated from the size of the pipes and from the measured discharge was found for the 2 in. supply to be 0·18 over a considerable range of discharge, while for the ½ in. supply the ratio was as low as 0·065. These large reductions in energy were unavoidable, for (apart from a small contribution due to shock at entry to the tank) they were caused by friction between the tank interior and the water, every particle of which completed many revolutions before being discharged through the trumpet.

5. COMPARISON OF THE THEORETICAL AND THE OBSERVED DISCHARGES

An approximate expression for the discharge in the presence of an air-core may be obtained on the assumption that at the throat of the trumpet the radial velocity is negligible, i.e. that the surface of the core is vertical. At this horizontal section the flow can be analyzed by Bernoulli's equation. If L is the difference of height between the crest and the throat, a and b are the radii of the throat and the core, and h' and

w are the pressure head and the downward velocity at radius r, it follows that for frictionless flow

$$h'+\frac{w^2+c^2/r^2}{2g}=H+L=H'. \tag{1}$$

Here the small term has been omitted which arises from the tangential kinetic energy at a large radius in the circular tank. The variation of h' over the annular cross-section is seen to be

$$h'=\frac{c^2}{2g}\left(\frac{1}{b^2}-\frac{1}{r^2}\right). \tag{2}$$

Strictly speaking, γ/b (where γ is the surface tension) should be subtracted from the right-hand side of (2), but investigation showed that this term was negligible for the relatively powerful vortices now under consideration. On combining (1) and (2) we find that w has the value, constant over the cross-section, given by

$$w=\left(2gH'-\frac{c^2}{b^2}\right)^{\frac{1}{2}}, \tag{3}$$

and that therefore the discharge Q is

$$Q=\pi a^2\left(1-\frac{b^2}{a^2}\right)\left(2gH'-\frac{c^2}{b^2}\right)^{\frac{1}{2}}. \tag{4}$$

Now for stability of flow b must adjust itself so that the discharge is maximum, and the condition $\partial Q/\partial b=0$ yields

$$b^2=\frac{1}{8gH'}\{c^2+(c^4+16a^2c^2gH')^{\frac{1}{2}}\}. \tag{5}$$

The solution containing a negative sign before the square root is clearly inadmissible. Thus (4) and (5) taken together provide an expression for Q when either c or b is given.

At Sir Geoffrey Taylor's suggestion an alternative proof of (5) is given below, based on the conception of Reynolds (1886) that over the critical cross-section the velocity of streaming is equal to that of a small wave in the fluid. This idea was applied originally to gas flow through a nozzle, where at the throat the streaming velocity was shown to be identical with the local velocity of sound. But it may also be used with liquid flow, and when employed to solve the problem of the broad-crested weir it gives at once the familiar result that the depth of the stream over the weir is two-thirds of the total head. In the present case it is necessary to calculate the velocity V of a 'long' wave of annular form moving in an axial direction on the rotating surface at the throat. If q is the additional velocity due to the wave motion at points where the infinitely small surface elevation is η, the equation of continuity is

$$\{\pi(a^2-b^2)+2\pi b\eta\}(V+q)=\pi(a^2-b^2)V. \tag{6}$$

Now the increase of pressure at any radius due to the spinning portion between radii b and $b-\eta$ is

$$\frac{c^2}{2}\left(\frac{1}{(b-\eta)^2}-\frac{1}{b^2}\right); \tag{7}$$

hence, ρ being the density and δp the excess of pressure owing to the wave motion,

$$\frac{\delta p}{\rho}+\frac{c^2}{2}\left(\frac{1}{(b-\eta)^2}-\frac{1}{b^2}\right)+\tfrac{1}{2}(V+q)^2=\frac{V^2}{2}. \tag{8}$$

On substitution from (6), (8) reduces to

$$\frac{\delta p}{\rho}+\frac{c^2\eta}{b^3}=\frac{2V^2b\eta}{a^2-b^2}, \tag{9}$$

η being taken as small. Thus the free surface conditon $\delta p = 0$ is satisfied when

$$V=\left\{\frac{c^2(a^2-b^2)}{2b^4}\right\}^{\frac{1}{2}}. \tag{10}$$

Finally, on identifying V with w and equating (3) and (10), we obtain (5).

The above conclusions hold good only when the throat controls the flow; if Q is determined by weir action, b can be found from (4) when c is specified. The experiments to be described were confined to the former conditions, and, as explained in § 4, it was not found feasible to measure the swirl close to the trumpet crest. A pressure tapping was therefore drilled in the throat of bellmouth B and connected by a tube of very small diameter to a water gauge outside the tank. Then from (2) a relation between b and the effective value of c could be obtained in the form

$$h=\frac{c^2}{2g}\left(\frac{1}{b^2}-\frac{1}{a^2}\right), \tag{11}$$

where h is the throat pressure head. The combination of this with (5) yields

$$\frac{b^2}{a^2}=\frac{1}{4H'}[2H'-h\pm\{(2H'-h)^2-8H'h\}^{\frac{1}{2}}] \tag{12}$$

and

$$\frac{c^2}{b^2}=\frac{g}{2}[2H'+h\pm\{(2H'-h)^2-8H'h\}^{\frac{1}{2}}], \tag{13}$$

in which expressions like signs in front of the square roots are to be taken with like. An ambiguity thus arises, for it is impossible from measurements of H' and h only to decide which of the two values of c and b are the true ones; the same value of h can be given by a stream possessing a relatively small swirl but large radial thickness and one in which the swirl is great and the radial thickness is small. The theoretical values of the discharge Q, to be compared with those observed, can be found from (4), (12) and (13), but the range over which the theory is valid is limited to $H'/h \geqslant 2{\cdot}91$ approximately. This restriction arises in (12) and (13), where the values of the expressions under the square roots must not fall below zero. The other limit $H'/h \leqslant 0{\cdot}09$, similarly obtained, is meaningless, for it leads to negative values of b^2/a^2.

Several sets of observations are displayed in table 1, to which have been added the corresponding double values of the theoretical discharges. It is obvious that the experiments were carried out in the range to which the negative signs apply, and that the theoretical discharges obtained with the positive signs may be ignored. From the last column in table 1 it appears that the observed and calculated discharges are in fair agreement in the range over which the theory is applicable. At low heads the theory overestimates the discharge; at high heads with weak swirls it predicts too low a value, the same tendency being noticeable when the limit $H'/h = 2{\cdot}91$ is approached, as is to be expected. Evidently under conditions near and below this critical value of H'/h the effects of friction within the trumpet become predominant, but a theory which takes these effects into account is not at present available. Swirling flow through a trumpet is one of the few cases where friction enhances the discharge. If the liquid suddenly became 'perfect', it is clear that the level in the tank would immediately rise owing to the reduction or even cessation of the discharge through the throat.

TABLE 1. COMPARISON OF OBSERVED AND THEORETICAL DISCHARGES.

Bellmouth *B*. Inch-second units.

observed				with positive signs	with negative signs	
H'	h	H'/h	Q	theoretical Q	theoretical Q	observed Q ÷ theoretical Q
Constant 1 in. tangential supply, 39·5 cu.in./sec.						
4·38	1·26	3·48	39·5	13·4	45·6	0·87
7·66	2·70	2·84	55·0	—	—	—
8·88	3·11	2·86	59·6	—	—	—
11·58	4·20	2·76	71·2	—	—	—
15·10	5·51	2·74	91·0	—	—	—
Pure 2 in. tangential supply						
4·04	0·98	4·12	40·9	8·7	50·0	0·82
4·51	1·19	3·79	43·4	11·0	50·0	0·87
5·28	1·50	3·52	47·2	14·2	50·7	0·93
6·58	2·05	3·21	51·8	20·5	50·4	1·03
6·95	2·15	3·23	53·2	20·6	52·4	1·02
7·94	2·62	3·03	56·0	27·7	49·0	1·14
8·00	2·64	3·03	56·0	27·8	49·2	1·14
9·04	3·05	2·96	60·3	33·2	48·2	1·25
10·28	3·56	2·89	63·9	—	—	—
11·49	4·08	2·82	67·6	—	—	—
Constant 1 in. tangential supply, 20·0 cu.in./sec.						
4·42	1·17	3·78	48·6	11·0	49·3	0·99
5·14	1·44	3·57	53·2	13·6	50·7	1·05
5·43	1·45	3·74	55·5	12·4	54·3	1·02
7·67	1·99	3·85	72·5	13·9	66·0	1·10
9·26	2·23	4·15	83·9	13·0	76·1	1·10
Constant 1 in. tangential supply, 14·7 cu.in./sec.						
4·23	0·87	4·86	54·1	6·4	55·5	0·97
5·89	1·11	5·31	68·7	6·4	67·7	1·01
7·90	1·46	5·41	83·0	7·2	78·8	1·05
10·96	1·80	6·09	102	6·8	96·2	1·06

Table 1 shows that within certain limits an estimate of the probable discharge may be made when H' and c are specified. Now (5) indicates the existence of an air-core however small c may be, yet trumpet tests gave results similar to figure 8, showing that flow without a complete air-core did in fact occur when the swirl was very low. Experiment confirmed that with dimples and very slender vortices surface tension was an important factor, for they were momentarily unsettled and sometimes destroyed by the introduction of a few drops of methylated spirit on the surface nearby. This treatment had no effect upon strong vortices. It was not, however, found possible to make measurements of feeble disturbances owing to their transient nature, thus an estimate cannot be given of the size of the smallest vortex which occurred. In the tests recorded in table 1 the least (calculated) value of b was in the last, where it was 0·20 in.

The existence of dimples (or embryonic vortices) situated at high heads above the trumpet is interesting, for at first sight the absence of an air-core below the dimple suggests infinite velocities there. The insertion of coloured water into the tank showed that even very close to the dimple the water was rotating very nearly in horizontal planes, but a narrow spinning tube of liquid descended with considerable velocity from the base of the dimple into the trumpet. Its diameter increased with depth as it entrained the surrounding water. Thus the motion was three-dimensional, and no question of infinite velocities arises. Gibson (1946) suggested that the motion could be explained by supposing the existence of a compound vortex (sometimes termed a Rankine combined vortex), which is a combination of a free vortex with a forced vortex inside it. This view seems untenable in the light of the observations described above.

Another attempt to support the theory was made by measuring the diameter of the air-core at the throat. The observations proved difficult owing to the corrugations on the water surface; they were carried out by fixing a short horizontal rod, pointed at both extremities, to the end of a long vertical bar which was traversed across the core. Two methods are available for calculating the theoretical values of b:

(i) By means of (12), employing the observations of H' and h.

(ii) On eliminating c from (4) and (5) we obtain

$$y^3 + Py - 2P = 0, \tag{14}$$

where $$y = 1 - b^2/a^2 \quad \text{and} \quad P = Q^2/(2gH'\pi^2 a^4).$$

Since P is positive this cubic equation has one positive real root, hence b can be determined from the observations of H' and Q.

The results of this investigation, given in table 2, confirm that the observations lay in the range to which the negative signs in (12) and (13) are applicable. The

TABLE 2. COMPARISON OF OBSERVED AND THEORETICAL VALUES OF b

Bellmouth B. Inch-second units.

					b calculated from		
					H' and h		
pure tangential supply	observed H'	h	Q	b	with positive signs	with negative signs	H' and Q
2 in.	3·62	0·95	39·4	0·32	0·53	0·28	0·32
1 in.	4·38	1·26	39·5	0·34	0·51	0·31	0·34
½ in.	4·25	1·33	34·6	$0{\cdot}31_6$	0·48	0·34	0·37

agreement between experiment and theory is fairly good except in the last set, which was taken with the powerful vortex produced by the ½ in. supply.

6. AIR MEASUREMENTS

The formation of an air-core led to no significant entrainment of air provided that the core was steady. The downward motion of the air within the core was due purely to surface friction and was arrested by the slightest rise in the pressure in the bell. Only when the vortex was feeble and irregular was there a definite entrainment of air insensitive to changes in the back pressure of the order of $\frac{1}{10}$ in. of water. Under these conditions a succession of air bubbles passed down the pipe, and a positive driving force was provided by the intervening water. This state of affairs ensued at low heads when surging set in and at high heads when the vortex spluttered, but the volume of air drawn down was never found to be large. With spluttering the maximum ratio of the air flow to the water flow was 0·28 for bellmouth *A* and 0·20 for bellmouth *B*, while for the surging condition the maximum ratio was much smaller, being only 0·04 for bellmouth *A*.

7. SURFACE OSCILLATIONS IN THE CIRCULAR TANK

The progress of the work was occasionally disturbed by the appearance of waves in the circular tank, which were of sufficient amplitude to impede the setting of the hook gauge. Sometimes the motion was seen to possess a nodal diameter rotating steadily in the tank, and hence the origin of the waves was traced to resonance between the natural periods of the water in the tank and of the disturbances (previously mentioned) in the vortex. This system can easily be excited in a tumbler, but no method of maintaining it indefinitely seems to have been previously noticed. Rayleigh (1876) and Lamb (1932) showed that on deep and stationary water in a circular tank waves of various kinds are possible. For the symmetrical type in a tank of radius R the period is $2\pi/(gk)^{\frac{1}{2}}$, where k is given by the roots of $J_0'(kR) = 0$, while when there are s rotating nodal diameters the period is $2\pi s/(gk)^{\frac{1}{2}}$, k being determined by the roots of $J_s'(kR) = 0$. From these results the following theoretical periods (in sec.) were calculated:

mode	1st	2nd	3rd	4th	5th
$s = 0$	0·90	0·66	0·55	0·48	0·43
$s = 1$	1·33	0·76	0·60	0·51	0·46

In the course of the experiments disturbances were noticed having periods of 1·32, 1·30 (twice), 0·91, 0·74 (thrice), 0·66 (twice), 0·59 (twice), 0·46, 0·43 sec. The shape of the single nodal diameter in the second mode was checked with hook gauges, and a node was found at $r/R = 0{\cdot}70$; the theoretical value is 0·719. The agreement between the two sets of figures shows that the existence of the air-core and of the streaming motion, which except near the centre was small compared with the wave velocities, had little effect upon the natural periods of the water. It is also clear that the disturbances in the vortex were not capricious but possessed definite frequencies.

We are greatly indebted to Sir Geoffrey Taylor for much helpful discussion of § 5.

REFERENCES

Binnie, A. M. 1938 *Proc. Roy. Soc.* A, **168**, 219.

Binnie, A. M. & Wright, R. K. 1941 *J. Inst. Civil Engrs*, **15**, 197.

Binnie, G. M. 1938 *J. Inst. Civil Engrs*, **10**, 65.

Gibson, A. H. 1911 *Mem. Manch. Lit. Phil. Soc.* **55**, no. 7.

Gibson, A. H. 1946 *Hydraulics and its applications*, p. 104. London: Constable.

Goldstein, S. (ed.) 1938 *Modern developments in fluid dynamics*, p. 104. Oxford: Clarendon Press.

Lamb, Sir H. 1932 *Hydrodynamics*, pp. 28, 287, 441. Cambridge University Press.

Rayleigh, Lord 1876 *Phil. Mag.* (5), **1**, 257 (*Papers*, **1**, 251).

Rayleigh, Lord 1892 *Phil. Mag.* (5), **34**, 177 (*Papers*, **3**, 594).

Reynolds, O. 1886 *Phil. Mag.* (5), **21**, 185 (*Papers*, **2**, 311).

Southwell, R. V. 1946 *Relaxation methods in theoretical physics*, ch. VI. Oxford: Clarendon Press. Also *Phil. Trans.* A, **240**, 117.

The kinetic salt effect in monolayer reactions

By J. T. Davies and E. K. Rideal, F.R.S.

Davy Faraday Laboratory, Royal Institution

(*Received* 22 *March* 1948)

The rate of hydrolysis of the ester group in the monocetylsuccinate ion by hydroxyl ion has been studied as a monolayer reaction at various alkali strengths and in the presence of neutral salts. The Donnan equilibrium equations have been employed to calculate the concentration of the hydroxyl ions near the ester linkage, and the hydrolysis constants are found to show a very marked decrease in divergence when referred to the concentration of alkali in the surface, rather than that in the bulk phase.

Activity coefficients of the ions in the surface phase have been calculated by the method of linear log plots of Harned, and Brönsted's theory of specific ion interaction was extended to treat the case of the surface phase in the presence of added neutral salt. The slowing up of the alkaline hydrolysis of a neutral ester of a long-chain acid can also be quantitatively explained assuming the existence of a 'surface phase', as can also the potential difference between the surface and the bulk.

Introduction

The work of Hughes & Rideal (1933) on the oxidation of unsaturated fatty acids, of Fosbinder & Rideal (1933) on the alkaline hydrolysis of γ-stearolactone, and of Alexander & Schulman (1937) and Alexander & Rideal (1937) on the alkaline hydrolysis of long-chain esters, has shown that monolayer reactions may differ from those in bulk in the possibility of control of the configuration of the long-chain molecule in the surface, with alteration in the steric factor and energy of activation of the reaction. These surface reactions have revealed that the velocity varies linearly with the concentration of reactant in the substrate on which the film is spread, indicating a proportionality between the amount of soluble reactant just under the film and in the bulk phase, and that the activity coefficients of reactants and transition complex need not be taken into account. Although this is true for a film of neutral molecules, a reaction between ionogenic layers, e.g. an ionic ester in the film and hydroxyl ions in the bulk phase, should involve a consideration of the redistribution of the soluble ions near the surface. Such has proved to be the case for the hydrolysis of a monolayer of monocetylsuccinate ions (Davies 1948).

We consider here this reaction more fully and describe a neutral salt effect resulting not only from a change in activity coefficients as in bulk reactions, but also from a change in the ionic concentrations in the monolayer. This has been deduced assuming a definite 'surface phase', and applying the equations originally developed by Donnan, and applied by Wilson (1916) and later by Danielli (1941).

Theoretical

Assuming that the rate of hydrolysis of the ester-carboxylate ion is determined by the concentration of an activated complex with hydroxyl ion,

$$-\frac{d(\bar{C}_{R'})}{dt} = k^0 \bar{C}_{R''}, \qquad (1)$$

where $\bar{C}_{R'}$ is the concentration in the 'surface phase' of monocetylsuccinate ions, t is the time, and $\bar{C}_{R''}$ is the concentration of activated complex in the surface phase. k^0 is a constant.

If an equilibrium is assumed between reactants and activated complex,

$$K = \frac{\bar{a}_{R''}}{\bar{a}_{\mathrm{OH}'}\,\bar{a}_{R'}}, \tag{2}$$

where K is a thermodynamic constant, and $\bar{a}$ refers to activity in the surface. This may be written as

$$K = \frac{(\bar{\gamma}_{\pm\mathrm{Na}_2R})^3\,\bar{C}_{R''}}{(\bar{\gamma}_{\pm\mathrm{NaOH}})^2\,(\bar{\gamma}_{\pm\mathrm{Na}R})^2\,\bar{C}_{\mathrm{OH}'}\,\bar{C}_{R'}}, \tag{3}$$

where $\bar{\gamma}_{\pm i}$ is the mean activity coefficient of species i in the surface. Substitution of $\bar{C}_{R''}$ from (3) into (1) leads to

$$-\frac{d(\bar{C}_{R'})}{dt} = k^0 K \bar{C}_{\mathrm{OH}'}\,\bar{C}_{R'}\,\frac{(\bar{\gamma}_{\pm\mathrm{NaOH}})^2\,(\bar{\gamma}_{\pm\mathrm{Na}R})^2}{(\bar{\gamma}_{\pm\mathrm{Na}_2R})^3}, \tag{4}$$

which may be abbreviated to

$$-\frac{d(\bar{C}_{R'})}{dt} = k^0 K \bar{C}_{\mathrm{OH}'}\,\bar{C}_{R'}\,F' \tag{5}$$

where F' is the activity factor.

Integrating (5),

$$-\int_1^2 \frac{d(\bar{C}_{R'})}{\bar{C}_{R'}} = \int_{t_1}^{t_2} k^0 K \bar{C}_{\mathrm{OH}'}\,F'\,dt. \tag{6}$$

To simplify the integration of the right-hand term, it is assumed that, as a first approximation, F' and $\bar{C}_{\mathrm{OH}'}$ are not time variable. Carrying out the integration under these conditions,

$$\ln(\bar{C}_{R'})_1 - \ln(\bar{C}_{R'})_2 = \bar{C}_{\mathrm{OH}'}\,k^0 K F'(t_2 - t_1). \tag{7}$$

Kk^0 can be written as a single constant, k_0, and, further, k may be written for $k_0 F'$. Thus

$$k = k_0 F' = K k^0 F'. \tag{8}$$

To obtain k_0, it is necessary to know k and F', the former depending on $\bar{C}_{\mathrm{OH}'}$, assumed constant over small time intervals. The values of the activity coefficients in the surface are necessary for the evaluation of F'; and $\bar{C}_{\mathrm{OH}'}$ implies a knowledge of the ionic redistribution at the surface. Davies (1948) has applied the Donnan equilibrium equations in the evaluation of $\bar{C}_{\mathrm{OH}'}$, but for an accurate analysis it is necessary to know the values of the activity coefficients in the bulk and in the surface for use in the Donnan equations

$$C_{\mathrm{Na}^\cdot}\,C_{\mathrm{OH}'}(\bar{\gamma}_{\pm\mathrm{NaOH}})^2 = \bar{C}_{\mathrm{Na}^\cdot}\,\bar{C}_{\mathrm{OH}'}(\bar{\gamma}_{\pm\mathrm{NaOH}})^2. \tag{9}$$

Assuming electrical neutrality of the surface phase

$$\bar{C}_{R'} + \bar{C}_{\mathrm{OH}} = \bar{C}_{\mathrm{Na}^\cdot}. \tag{10}$$

The concentration of surface-active ions in the surface layer is given by

$$\bar{C}_{R'} = \frac{10^{24}}{Ax} \times \frac{1000}{N}, \tag{11}$$

where N is the Avogadro number, A is the area occupied by the long-chain ions, and x is the thickness of the surface layer, expressed in Å. The previous value of 10 Å for x (Davies 1948) will be used throughout.

If the substrate contains an added neutral salt, say NaCl, equation (10) becomes

$$\bar{C}_{R'} + \bar{C}_{\mathrm{OH'}} + \bar{C}_{\mathrm{Cl'}} = \bar{C}_{\mathrm{Na^{\cdot}}}, \tag{12}$$

and a further equation may be formulated

$$C_{\mathrm{Na^{\cdot}}} C_{\mathrm{Cl'}} (\gamma_{\pm \mathrm{NaCl}})^2 = \bar{C}_{\mathrm{Na^{\cdot}}} \bar{C}_{\mathrm{Cl'}} (\bar{\gamma}_{\pm \mathrm{NaCl}})^2. \tag{13}$$

It is thus possible to solve the equations completely for $\bar{C}_{\mathrm{OH'}}$, required in (7), if activity coefficients are known. This applies both in the absence and presence of added neutral salt. The concentrations in the bulk are effectively unchanged by the adsorption at the surface.

The evaluation of the activity coefficients required both in the accurate Donnan equations and in the factor F' presents several difficulties.

(*a*) A preliminary value for the ionic strength in the surface phase must be known.

(*b*) The surface phase is too concentrated (about 3 to 6 N) for any simple equation of the Debye-Hückel type to hold.

(*c*) The surface phase contains two or more electrolytes. Comparatively little work has been done on activity coefficients of one electrolyte in the presence of another in concentrated solutions, and no figures were known for the activity coefficients of an ion pair in the presence of two other ion pairs with a common cation.

(*d*) Assumptions have to be made about the activity coefficient in the surface of the insoluble monocetylsuccinate ion.

(*e*) The evaluation of $\bar{\gamma}_{\pm \mathrm{Na}_2 R}$ is complicated by the separation of the ionic groups by two saturated carbon atoms, the paucity of knowledge concerning the activity coefficients of doubly charged ions, and the uncertainty concerning the ionic environment of the hydroxyl part of the complex in the surface. It has been found that at the present stage $\bar{\gamma}_{\pm \mathrm{Na}_2 R}$ cannot be evaluated directly, and only indirect methods can be used to find its variation at different surface ionic strengths.

Consider now the simplest case, that of sodium cetylsuccinate and NaOH in the surface phase, and NaOH alone in the bulk. In the latter the concentration $C_{\mathrm{Na^{\cdot}}}$ or $C_{\mathrm{OH'}}$, is known, as is also the mean activity coefficient, but before the accurate Donnan equations can be solved, it is necessary to know $\bar{\gamma}_{\pm \mathrm{NaOH}}$, a quantity also required in the evaluation of the kinetic activity factor F'.

The method developed for the estimation of $\bar{\gamma}_{\pm \mathrm{NaOH}}$ depends on the linearity, as shown by Harned & Owen (1943*a*), of the logarithmic plots of the mean activity coefficients of a uni-univalent salt in the presence of another uni-univalent ion pair with a common ion, against composition at constant total ionic strength.

The determination of $\bar{\gamma}_{\pm \mathrm{Na}R}$ presents some difficulty, and it is necessary to assume that the activity coefficients of sodium cetylsuccinate in the presence of other ions, and those of soluble ions in the presence of the long-chain ion can be evaluated using the values for a system in which NaR is replaced by NaI in equivalent concentration. The justification for this lies in the structural similarity of that part of the monocetylsuccinate ion in the surface phase to sodium acetate, the activity coefficients of the latter being very close, over a wide range of concentrations in aqueous solution to those of NaI (Robinson 1935). This last step could have been eliminated had appropriate electrical measurements been made in systems containing sodium hydroxide and sodium acetate.

In evaluating $\bar{\gamma}_{\pm NaOH}$ in the absence of added salt, the ionic strength and composition of the surface phase may be determined to a first approximation by omitting activity coefficients from equation (9), and a plot of $\log \bar{\gamma}_{\pm NaOH}$ at constant ionic strength equal to that in the surface can be drawn using the values of Harned & Owen (1943*a*) for the activity coefficient of NaOH in very dilute solution in the presence of NaI at the ionic strength equal to that found in the surface. From the plot can be read the value of $\log \bar{\gamma}_{\pm NaOH}$ at the particular surface composition deduced from the simplified equations, and this activity coefficient, together with that of NaOH in the bulk, which is known (Harned & Owen 1943*b*), can be substituted into (9) and a more accurate value of surface ionic strength and composition results. The process can be repeated several times until no further change in the surface concentrations and activity coefficients is effected by resubstitution.

When a 'neutral' salt, such as NaCl, is added to the substrate, the surface contains chloride, hydroxyl, monocetylsuccinate and sodium ions. To treat this case, it was found necessary to extend the Brönsted (1922) theory of specific ion interactions, which agrees with the linear logarithmic plots of Harned & Owen (1943*a*). The full treatment of this theory for the case of two electrolytes with a common ion is given by these authors, and need not be given here. Application of the theory to the present case of varying amounts of one anion and a constant ratio of two others at a constant concentration of cation indicates that the linear relationship of the logarithms of the activity coefficient of the first species in the mixture still holds. Use is further made of the theory to evaluate such terms as the activity coefficient of NaCl in NaOH of strength M when the chloride is very dilute. This has been shown by Harned & Owen to be the same as the activity coefficient of very dilute NaOH in the presence of NaCl of strength M. To find the activity coefficient of NaCl in the surface, required in (13), the logarithms of the activity coefficients of this, in very dilute solution, in the presence of NaOH and of NaI, each at the approximate surface ionic strength, are used as limiting values in a linear plot of $\log \bar{\gamma}_{\pm NaCl}$ against molal concentration of NaI or NaOH at constant strength. The linearity of this plot can be shown to follow directly from the Brönsted theory. The approximate values of the surface ionic strength and composition are obtained from solution of (9), (12) and (13) assuming activity coefficients to be unity. The abscissa corresponding to the ratio of hydroxide to iodide (representing insoluble ion) in the logarithmic graph is thus known, and hence $\log \bar{\gamma}_{\pm NaCl}$ at the surface ionic strength in the presence of NaOH and NaI in the ratios, but not the concentrations, found in the surface, since NaCl is taken as extremely dilute in this plot (0·01 M).

A second plot is now constructed with $\log \bar{\gamma}_{\pm NaCl}$ as limiting values in the presence of NaCl and a constant ratio mixture of NaOH and NaI, again at the surface ionic strength. The abscissa corresponding to the amount of NaCl actually found in the surface gives the required activity coefficient. The value of $\bar{\gamma}_{\pm NaOH}$

is also required for an accurate solution of (9) and (13), and is obtained in the same way. The whole process must be repeated several times.

The neutral salts added were NaCl and NaI. It was hoped to use NaCNS, but the strong alkali decomposed this salt too rapidly. The strongly surface-active nature of these large anions would be found to affect the activity coefficients in the surface phase in a specific manner if there were any considerable effects due to adsorption. These were not found for NaI, however, which would indicate a predominance of normal ionic forces.

EXPERIMENTAL

The techniques already described by previous workers were used in this work. A Langmuir trough, in which the film pressures could be measured with an accuracy of 0·2 dyne/cm.$^{-1}$ and thermostated to 0·2° C, was employed. All reactions were carried out at constant surface pressure, and the film was confined by means of thin platinum ribbons.

Sodium hydroxide and sodium chloride used were A.R. materials. Pure sodium iodide was used, and all water and organic solvents were redistilled from an all-glass pyrex apparatus. Benzene was found to be a satisfactory solvent for the monocetylsuccinate.

Monocetylsuccinate was synthesized by Dr J. W. H. Oldham. Equimolecular amounts of succinic anhydride and cetyl alcohol were heated at 180° C for 2 hr. without a solvent. The resultant product was almost completely soluble in a solution of $KHCO_3$ in distilled water, on account of the carboxyl group present. The solution was then extracted with benzene to remove unreacted cetyl alcohol and di-ester, and acidified with HCl. Extracting with ether removed the monocetylsuccinate, which was washed, while still in ethereal solution, twice with distilled water to remove traces of HCl, and was finally dried by evaporation under reduced pressure on a water-bath. The remaining material was recrystallized from petrol ether till a constant melting-point product resulted, with m.p. of 61 to 63° C (cetyl alcohol 49·3° C). Titration with standard NaOH gave a mol.wt. of 340 (calculated 342), and alkaline hydrolysis gave rise to a further 0·83 equivalent of acid, a figure regarded as satisfactory in view of the difficulty in titrating long-chain substances. Cold aqueous ethanol was the solvent for these titrations.

All glassware was cleaned in hot chromic acid, followed by phosphoric acid to remove adsorbed chromium ions.

EVALUATION OF THE RATE CONSTANT

Equation (7) shows that for the rate constant k to be found, not only is $\bar{C}_{OH'}$ required, but also the rate of variation of the amount of long-chain ion in the surface. Previous workers (Alexander & Rideal 1937) have found that the surface moment is linearly related to the amount of reactant in the surface in a reaction at constant pressure. In the present work, however, this simple relationship did not hold, owing, no doubt, to a dipole interaction of the type described by Marsden & Schulman (1938), who found that films containing an ion and a dipole often show large interaction effects. Figure 1 shows the force-area and potential and surface-moment curves for monocetylsuccinate on N alkali. Figure 2 shows the ΔV area and $\Delta V.A$ values for mixed films, where A refers to the area per molecule and ΔV is the surface potential. The functions are seen not to be linear except for the initial portion of the area curve, between pure ester-acid and a mixture with about 30 % alcohol. The large interaction at points corresponding to appreciable amounts of alcohol in the acid-ester film is to be correlated with the existence of three dipoles in the surface, one of them carrying a negative charge.

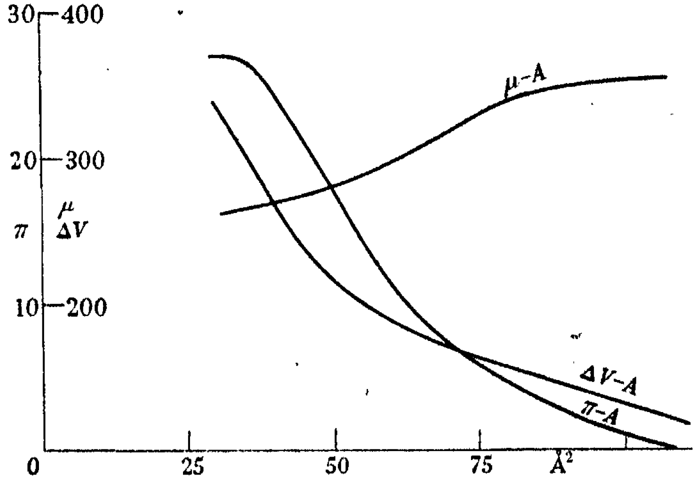


FIGURE 1. Monocetylsuccinate on alkali. pH 14·02.

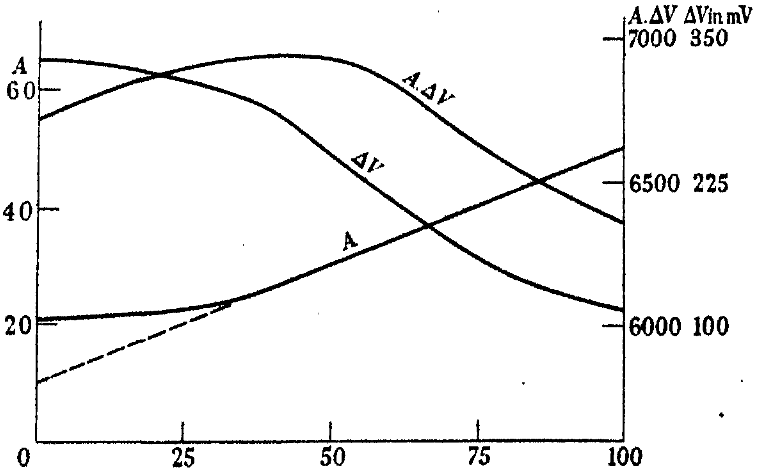


FIGURE 2

The method adopted for the measurement of the time variation of $\log(\bar{C}_{R'})$ depended on the linearity of the plot of area against composition in the initial part of the reaction, and equation (7) was used in the form

$$\ln(A_1 - 10) - \ln(A_2 - 10) = k.\bar{C}_{OH'}(t_2 - t_1), \tag{14}$$

where 10 Å^2 is the area per molecule the alcohol film would exhibit if the linear portion of the curve persisted to 100 % cetyl alcohol. This plot of ln $(A-10)$ against time gave satisfactory straight lines. Only the first 15 % or so of the reaction was studied, since during the hydrolysis $\bar{C}_{R'}$ changes, and with it F' and $\bar{C}_{\text{OH}'}$. The temperature of hydrolysis was in all cases close to 25° C, and any corrections to exactly 25·0° C were made, using the energy of activation of the reaction. The rate constants, k', calculated on the basis of a bulk concentration of NaOH of 1·044 N, and expressed as min.$^{-1}$ mole^{-1} litres were found to vary with the temperature and pressure as shown below.

HYDROLYSIS CONSTANTS ON 1·044 NaOH

$\Pi = 10$ dynes/cm.

T° C	21·5	25·8	31·1
k'	$1{\cdot}08\times10^{-2}$	$1{\cdot}61\times10^{-2}$	$2{\cdot}66\times10^{-2}$

$E = 16{,}600$ cal./g.mol.

$\Pi = 3$ dynes/cm.

T° C	21·5	24·8	30·2
k'	$0{\cdot}590\times10^{-2}$	$0{\cdot}845\times10^{-2}$	$1{\cdot}37\times10^{-2}$

$E = 17{,}000$ cal./g.mol.

RESULTS

In figure 3 is plotted the rate $(k'\times10^3)\times C_{\text{OH}'}$ where k' is the hydrolysis constant calculated directly on the basis of a bulk concentration of hydroxyl ion, against this last quantity. It is clear that the rate is not linearly related to the amount of alkali in the bulk, even in this case where no neutral salts have been added.

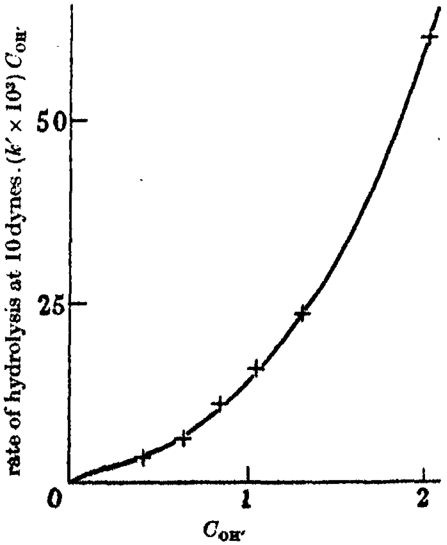


FIGURE 3

Table 1 gives the apparent hydrolysis constants, k', at various concentrations of alkali and neutral salt in the bulk. The figures for the hydrolysis with alkali alone are those given previously (Davies 1948). Activity coefficients were calculated by the methods described above, and are shown in the sixth and eighth columns of the table. In the seventh are the activity coefficients of NaOH in the bulk phase for purposes of comparison. They are taken from the table given by Harned & Owen (1943 *b*).

TABLE 1

$C_{OH'}$	$C_{Cl'}$	$C_{I'}$	$\bar{C}_{R'}$	$\bar{C}_{OH'}$	$\bar{\gamma}_{\pm NaOH}$	$\gamma_{\pm NaOH}$	$\bar{\gamma}_{\pm NaR}$	$k' \times 10^3$	$k \times 10^3$	surface ionic strength
0·428	—	—	3·40	0·076	0·59	0·70	1·04	7·9	44·8	3·48
0·646	—	—	3·25	0·16	0·59	0·68	1·02	9·0	35·7	3·41
0·855	—	—	3·15	0·28	0·60	0·68	1·00	12·8	39·4	3·43
1·044	—	—	3·10	0·40	0·60	0·68	1·00	15·0	38·9	3·50
1·303	—	—	3·08	0·64	0·61	0·70	0·98	17·8	36·2	3·72
1·970	—	—	3·07	1·11	0·64	0·72	0·98	31·0	55·0	4·18
0·855	1·41	—	3·04	0·55	0·62	0·66	1·06	17·7	27·8	4·00
0·855	2·83	—	3·02	0·68	0·66	0·67	1·24	22·2	28·0	4·79
0·854	4·24	—	3·02	0·75	0·73	0·71	1·55	29·8	34·1	5·62
1·112	1·41	—	2·99	0·75	0·64	0·67	1·07	23·0	34·0	4·19
1·126	—	1·25	2·97	0·71	0·63	0·65	1·02	19·6	31·2	4·02
1·126	—	2·50	2·84	0·87	0·64	0·66	1·15	21·0	27·1	4·84

k values in min.$^{-1}$ mole^{-1} litres; all films at 10 dynes cm.$^{-1}$ and 25° C.

The concentration of hydroxyl ion in the surface phase, $\bar{C}_{OH'}$ has been obtained by successive approximations. As an example of this procedure, the case of hydrolysis of the ester ion on 0·855 N alkali containing 1·41 N-NaCl will be considered. Since the thickness of the surface phase is 10 Å, and the film occupies 54·7 Å per molecule, the strength of the surface phase, $\bar{C}_{R'}$, is given by

$$\frac{10^{24} \times 10^3}{6{\cdot}02 \times 10^{23} \times 10 \times 54{\cdot}7} = 3{\cdot}04 \text{ mol./litre.}$$

Hence (12) gives $\qquad \bar{C}_{OH'} + 3{\cdot}04 + \bar{C}_{Cl'} = \bar{C}_{Na^{\cdot}}.$

and (9) and (13) may be written

$$\bar{C}_{Na^{\cdot}} . \bar{C}_{OH'} = (2{\cdot}27)\ (0{\cdot}855), \quad \bar{C}_{Na^{\cdot}} . \bar{C}_{Cl'} = (2{\cdot}27)\ (1{\cdot}41),$$

where activity coefficients have been assumed equal as a first approximation. These three simultaneous equations on solution give

$$\bar{C}_{Na^{\cdot}} = 4{\cdot}25, \quad \bar{C}_{OH'} = 0{\cdot}457, \quad \bar{C}_{Cl'} = 0{\cdot}753.$$

From the experimental figures given by Harned & Owen (1943 *a*) and the equations of Brönsted (omitting salting-out coefficients),

$$\gamma_{\pm NaOH\ (NaCl)} = \gamma_{\pm NaCl,\ (NaOH)} = 0{\cdot}637.$$

where an ionic strength of 4·25 is assumed, and where $\gamma_{\pm NaOH\,(NaCl)}$ means the activity coefficient of very dilute NaOH in the presence of NaCl of the particular ionic strength.

From the same sources

$$\ln \gamma_{\pm \mathrm{NaCl}(\mathrm{Na}R)} = \tfrac{1}{2}(\ln \gamma_{\pm \mathrm{NaCl}} + \ln \gamma_{\pm \mathrm{Na}R}),$$

which gives, assuming the similarity of behaviour of NaR and NaI, as explained above,

$$\gamma_{\pm \mathrm{NaCl}(\mathrm{Na}R)} = 0{\cdot}989.$$

The fraction of NaR in a hypothetical phase of strength 4·25 N, containing only NaR and NaOH in the same proportions as those found as a first approximation in the surface phase is

$$\frac{3{\cdot}04}{3{\cdot}04 + 0{\cdot}457} = 0{\cdot}869,$$

and, using this figure, the activity coefficient of NaCl in a mixture of NaR and NaOH in the same ratio and at the same total strength as that in the surface can be found, since extension of the Brönsted theory has shown that the plot of $\ln \gamma_{\pm \mathrm{NaCl}}$ in very low concentrations against the proportion of R' or OH$'$ at constant total ionic strength is linear.

Another linear plot is now constructed in which the figure thus obtained for the activity coefficient of NaCl in the NaOH/NaR mixture and that of pure 4·25 N-NaCl are limiting ordinates, and the abscissa is the fraction of NaCl in the mixture. The final value of the activity coefficient in the surface is read off at the abscissa corresponding to the fraction of NaCl found as a first approximation, i.e.

$$0{\cdot}753/4{\cdot}25 = 0{\cdot}177,$$

and $\gamma_{\pm \mathrm{NaCl}}$ is thus found to be 0·91.

By similar processes, $\gamma_{\pm \mathrm{NaOH}}$ and $\gamma_{\pm \mathrm{Na}R}$ in the surface are found to be 0·62 and 1·06, and the activity coefficients of the NaOH and NaCl in the presence of each other in the bulk phase are 0·656 and 0·656, the agreement being coincidental.

Equations (9) and (13) may now be formulated

$$\bar{C}_{\mathrm{Na}^{\cdot}} . \bar{C}_{\mathrm{OH}'} = (2{\cdot}27)\ (0{\cdot}855)\ \left(\frac{0{\cdot}66}{0{\cdot}62}\right)^2,$$

$$\bar{C}_{\mathrm{Na}^{\cdot}} . \bar{C}_{\mathrm{Cl}'} = (2{\cdot}27)\ (1{\cdot}41)\ \left(\frac{0{\cdot}66}{0{\cdot}91}\right)^2,$$

and, together with $$\bar{C}_{\mathrm{OH}'} + 3{\cdot}04 + \bar{C}_{\mathrm{Cl}'} = \bar{C}_{\mathrm{Na}^{\cdot}},$$

give $$\bar{C}_{\mathrm{Na}^{\cdot}} = 4{\cdot}01, \quad \bar{C}_{\mathrm{OH}'} = 0{\cdot}546, \quad \bar{C}_{\mathrm{Cl}'} = 0{\cdot}414.$$

No further approximations are necessary in this example, since the concentrations and activity coefficients are fairly low. In the more concentrated solutions, and in the case of added NaI, a further substitution and resolution is necessary. The use of activity coefficients thus increases $\bar{C}_{\mathrm{OH}'}$ from 0·46 to 0·55, with decrease in the velocity constant from $33{\cdot}1 \times 10^{-3}$ to $27{\cdot}8 \times 10^{-3}$.

Recalculations of the hydrolysis constant using the final values of $\bar{C}_{\mathrm{OH}'}$ gives the figures in the column headed $k \times 10^3$.

It was shown previously (Davies 1948) that use of $\bar{C}_{\mathrm{OH}'}$, calculated assuming activity coefficients to be unity, reduced the variation in the hydrolysis constant with concentration of alkali from 300 to 36 %, and table 1 shows that the application of the exact Donnan equations still leaves a variation of this order of magnitude. There is still, however, the effect of the kinetic activity factor F' to take into account. The activity coefficient of the doubly negatively charged intermediate complex is not known, but if a constant value of k_0 in equation (8) is assumed, it is possible to find the variation of this activity coefficient with composition and with ionic strength of the surface phase, since the activity coefficients of NaOH and the sodium cetylsuccinate in the surface are known. From (4) and (5)

$$F' = \frac{(\bar{\gamma}_{\pm \mathrm{NaOH}})^2 (\bar{\gamma}_{\pm \mathrm{Na}R})^2}{(\bar{\gamma}_{\pm \mathrm{Na}_2R})^3}, \tag{15}$$

and, from (8),

$$k_0 = \frac{k}{F'}$$

$$= \frac{k(\bar{\gamma}_{\pm \mathrm{Na}_2R})^3}{(\bar{\gamma}_{\pm \mathrm{NaOH}})^2 (\bar{\gamma}_{\pm \mathrm{Na}R})^2}. \tag{16}$$

Therefore

$$\frac{(\bar{\gamma}_{\pm \mathrm{Na}_2R})^3}{k_0} = \frac{(\bar{\gamma}_{\pm \mathrm{NaOH}})^2 (\bar{\gamma}_{\pm \mathrm{Na}R})^2}{k}$$

or

$$\frac{\bar{\gamma}_{\pm \mathrm{Na}_2R}}{(k_0)^{\frac{1}{3}}} = \left(\frac{(\bar{\gamma}_{\pm \mathrm{NaOH}})^2 (\bar{\gamma}_{\pm \mathrm{Na}R})^2}{k}\right)^{\frac{1}{3}}. \tag{17}$$

From this the variation of $\bar{\gamma}_{\pm \mathrm{Na}_2R}$ follows at once.

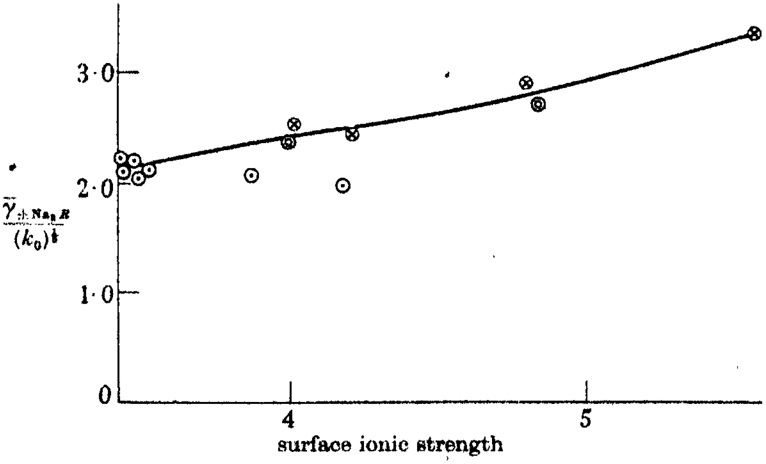


FIGURE 4. ⊙ NaOH alone, ⊗ NaOH + NaCl, ◉ NaOH + NaI.

Using the figures in table 1 to evaluate the right-hand member, figure 4 shows a plot of $\bar{\gamma}_{\pm \mathrm{Na}_2R}/(k_0)^{\frac{1}{3}}$ against surface ionic strength. It is remarkable that for the hydrolyses on alkali of strength 0·4 to 1·3 N the activity coefficients and ionic strength in the surface are so close together. The latter effect arises from a slight expansion of the monolayer on the more strongly alkaline substrates. Evidently NaCl and NaI affect the activity coefficient in similar ways, with no unusual effects from the large iodide ion, but NaOH in very strong solution behaves differently, and reduces

$\bar{\gamma}_{\pm Na_2R}$. Figure 5 shows a plot of $pH_B - pH_S$, subscripts referring to bulk and surface phases, against pH_B. Although at lower pH_B values the difference may be quite considerable, increase in alkali concentration greatly reduces the effect.

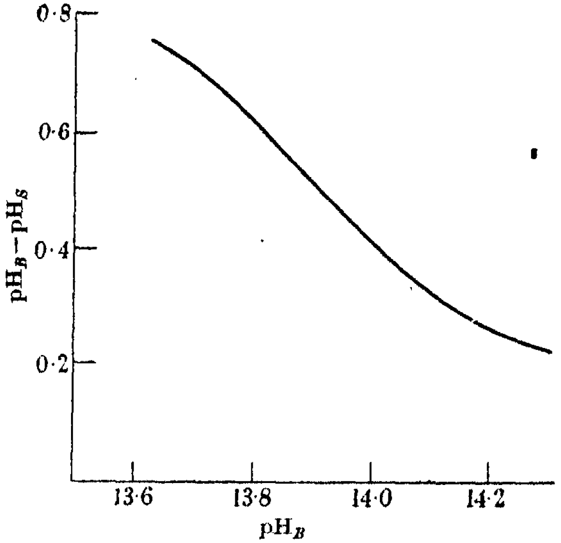


FIGURE 5

DISCUSSION

Of the various methods available for treating the distribution of ions at a charged interface, the Donnan distribution method is the most readily applicable to the present case, when ionization in the monolayer is practically complete. For kinetic equations, concentrations are required rather than chemical potentials, and measurements of the electrokinetic potential of a monolayer are difficult.

The Donnan equations, however, prove reasonably satisfactory for the present purpose, though there are various inherent sources of error in the assumptions and equations used in the above treatment.

These include those mentioned previously, of a constant value of $\bar{C}_{OH'}$ and F' over the part of the reaction in which measurements are made. This will be true only over very small intervals of time, but in practice only the first few minutes of the reaction were taken into account, so the effect is not serious. A further source of error is that of the electrical neutrality equation implying a great thickness of surface phase so that all the counter ions are taken into account, while a small value is necessary for the concentration of hydrogen ions to be that close to the surface.

An experimental difficulty is that the surface is more concentrated than some of the solutions used by Harned and co-workers in the evaluation of activity coefficients, and in these cases an extrapolation of the available curves is necessary.

The application of the Donnan treatment is not limited to cases of reaction between an ion and another ion in the surface. If, for example, a reactant molecule in the film is near a separate long-chain ion, there will still be a redistribution of ions, affecting the rate of hydrolysis. Such a case was encountered by Alexander &

Rideal (1937) in the hydrolysis of the neutral ester of a long-chain acid on alkali. The hydrolysis constant falls off as the reaction proceeds, and these authors showed, by using synthetic mixtures of long-chain acid and ester, that the effect was not due to contamination in the film, nor to the formation of a complex in the surface. They suggested that the cause of the slowing-up lay in the powerful electric field set up in the surface by the negatively charged carboxyl groups from the acid liberated during the hydrolysis. It was this field that reduced the rate of upward diffusion of the hydroxyl ions, slowing up the reaction. They thus assumed that a potential barrier, represented by E_E, was responsible for the slowing-up, so that E_E must increase during the reaction. For the hydrolysis of ethyl palmitate they found that the initial rate constant of 0·038 min.$^{-1}$ decreased to 0·023 min.$^{-1}$ after 25 % reaction, and to 0·013 min.$^{-1}$ after 50 % reaction. E_E was thus found empirically to be 325 and 630 calories.

Application of the Donnan equations, however, makes possible a theoretical calculation of the slowing-up factor and of the empirical energy barrier. For the case of 25 % reaction, for example, if the thickness of the surface layer is taken as one quarter of the distance between the centres of neighbouring ions in the surface, in this case about 6·5 Å, and if successive approximations are employed to obtain activity coefficients, as before, the rate constant is found to be reduced by a factor of 0·64, to 0·024 min.$^{-1}$, in good agreement with the observed figure of 0·023 min.$^{-1}$. The exact surface thickness does not affect this result greatly. If 4 Å, for example, had been taken instead of 6·5 Å, the rate constant would have been 0·018 min.$^{-1}$.

For the case of 50 % reactions similar considerations give the constant a value of 0·014 min.$^{-1}$, and the observed value is 0·013 min.$^{-1}$. The potential difference between the bulk and the surface, postulated by Alexander & Rideal, is the Donnan potential, which agrees with the values deduced empirically from the kinetic data.

The authors wish to thank Dr J. W. H. Oldham for the preparation of pure monocetylsuccinate, and one of us (J.T.D.) is indebted to the Council of Canterbury University College, New Zealand, for an award of the Lord Rutherford Memorial Research Fellowship.

References

Alexander, A. E. & Rideal, E. K. 1937 *Proc. Roy. Soc.* A, **163**, 70.
Alexander, A. E. & Schulman, J. H. 1937 *Proc. Roy. Soc.* A, **161**, 115.
Brönsted, J. N. 1922 Quoted by Harned & Owen (1943*a*).
Danielli, J. F. 1941 *Biochem. J.* **35**, 470.
Davies, J. T. 1948 *Report of Faraday Society, Bordeaux Conference* (in the Press).
Fosbinder, R. J. & Rideal, E. K. 1933 *Proc. Roy. Soc.* A, **143**, 61.
Harned, H. S. & Owen, B. B. 1943*a* *The physical chemistry of electrolytic solutions*, Chapter 14. New York: Reinhold Publishing Corporation.
Harned, H. S. & Owen, B. B. 1943*b* *The physical chemistry of electrolytic solutions*, p. 560. New York: Reinhold Publishing Corporation.
Hughes, A. H. & Rideal, E. K. 1933 *Proc. Roy. Soc.* A, **140**, 253.
Marsden, J. & Schulman, J. H. 1938 *Trans. Faraday Soc.* **34**, 748.
Robinson, R. A. 1935 *J. Amer. Chem. Soc.* **57**, 1161, 1165.
Wilson, J. A. 1916 *J. Amer. Chem. Soc.* **38**, 1982.

Atmospheric tides in the ionosphere
III. Lunar tidal variations at Canberra

By D. F. Martyn, D.Sc.
Australian Council for Scientific and Industrial Research

(*Communicated by Sir Edward Appleton, F.R.S.—Received* 16 *October* 1947)

A study has been made of the lunar variations in heights and critical frequencies of the E, F_1 and F_2 regions above Canberra, Australia. Semi-diurnal lunar variations have been found in all these quantities, and the harmonic coefficients determined. From the study of these it is concluded: (*a*) that ionospheric lunar variations are caused by ionic drift under the action of the 'dynamo' electric forces, and not by simple tidal rise and fall of isobaric surfaces; (*b*) that the lunar magnetic variations are not produced in the E, F_1 or F_2 regions.

1. Introduction

In two previous papers some consequences of the solar and lunar tidal motions in the ionosphere have been discussed. In the first of these (Martyn 1947*a*), afterwards referred to as I, a theory was propounded that major anomalies in the behaviour of the F_2 region were due to the action of the solar atmospheric tide. The theory showed that this tide, when proper account was taken of the earth's magnetic field, could cause notable *vertical* distortion of an ionized region. In the second paper (Martyn 1947*b*), afterwards referred to as II, F region data from the Huancayo Observatory were analyzed for lunar tidal variations. Large tidal variations in both the maximum electron density N(max.) and h (max.), the height at which this density occurs, were found to occur in the ionosphere at this site, which is near to the magnetic equator. It was also found that the phases and amplitudes of these lunar variations depended on solar time in a complex manner. The present paper describes the results of an investigation of the tidal variations at a moderate latitude (35° S.), where the theory indicates that the effects should be smaller in magnitude. The theory also indicates that the ionosphere, in its response to tidal motions, should behave like a non-linear system, so that the smaller tidal motions anticipated at Canberra should give rise to simpler ionospheric variations than were found at Huancayo. It is found in the course of the present work that these anticipations are borne out, and in consequence it has been possible to draw certain conclusions regarding the global lunar current circulations in the ionosphere.

2. Lunar tidal variations derived from ionospheric data

The data available consist of nine years of hourly measurements of the heights and critical penetration frequencies f^0 of the E, F_1 and F_2 regions at Canberra (lat. 35·3° S., long. 149° E.). They were made available by the kind permission of the Commonwealth Astronomer and the Radio Research Board, who were jointly responsible for the observational programme. The methods used in deriving the lunar variations have already been described in II. The reduction of the half million or so hourly readings involved is necessarily a laborious process. In the case of regions E and F_1 it has been necessary to use nearly all the available data in order to derive the lunar variations satisfactorily. This has been necessary partly because these regions can only be observed in daylight hours, and partly because of the smallness of the variations. In the case of region F_2 a satisfactory determination of the lunar variations has been possible using only three years of the available data. At the present stage it has seemed desirable to use the available computing resources to make a conspectus of the variations with reasonable accuracy at a number of different latitudes (and longitudes), rather than to obtain the maximum accuracy at any one site by the use of all available data at that site.

2·1. *Region E*

Hourly measurements of the minimum equivalent height h' of the E region, for the four years from 1941 to 1944, were analyzed for lunar variations, using the methods described in II. No significant tidal variations were found in the first, third or fourth harmonics. The second harmonic showed a well-marked variation.

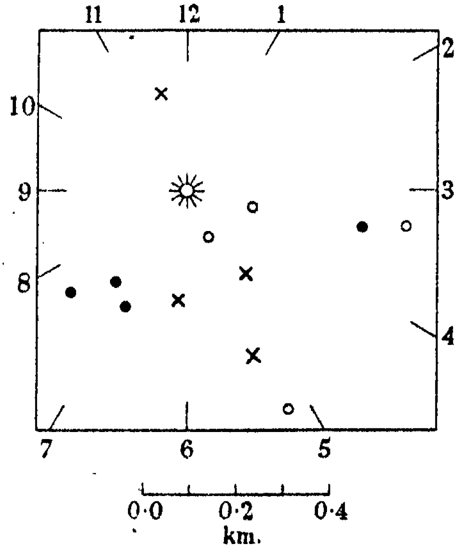


FIGURE 1. Harmonic dial for semi-diurnal lunar variation in h'_E at Canberra, 1941 to 1944. ● Months of northern solstice. ○ Months of southern solstice. × Equinoctial months.

For purposes of presentation the data have been treated in three seasonal periods, comprising the months of northern solstice (N), the equinoctial months (E), and those of southern solstice (S). In each year the harmonic coefficients P_2, t_2 (the amplitude in km., and the time of maximum amplitude in lunar hours reckoned from local transit) are displayed in the usual manner on the 12-hourly harmonic dial. The results for (N) months are plotted in figure 1 as filled in circles, those for (E) months as crosses, and those for (S) months by open circles.

It is seen that the points tend to group themselves about the lunar hour 5, showing the reality of a tidal variation in h'_E with maximum phase at this time. There is also a tendency for this phase to retard in passing from (S) months through (E) to (N) months. This is shown in table 1, which gives the average harmonic coefficients for all years for each season. The mean harmonic coefficients for all years and all seasons are $P_2 = 0{\cdot}187$ km. and $t_2 = 5{\cdot}1$ hr.

TABLE 1

season	P_2 (km.) for h'_E	t_2 (lunar hours)
N	0·19	6·3
E	0·16	5·3
S	0·28	4·2

The same procedures have been applied to the hourly values of critical penetration frequency f^0_E, with the results shown on the harmonic dial in figure 2. In this case, owing to the smallness of the variation, it has been necessary to use 10 years' data, covering the period 1937 to 1946. Again there is evidence of a lunar 12-hourly variation with maximum phase about 4 hr. In the case of critical frequencies, for reasons explained in II, it is desirable to plot the ratio of mean lunar variation in f^0 to the mean value of f^0 for the period considered. The quantity actually plotted is $200P_2/\text{mean}\,f^0_E$, which is the measure of the percentage variation in the maximum electron density N_m. As in the case of h'_E there is evidence of a seasonal retardation in the phase of the f^0_E lunar variation. The mean coefficients for each season are set out in table 2. The mean coefficients for all seasons over the 10 years are

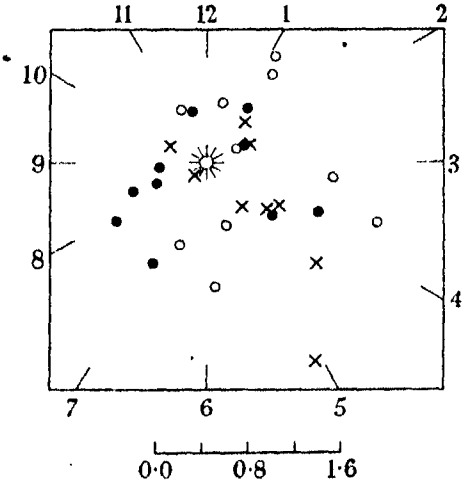


FIGURE 2. Harmonic dial for semi-diurnal lunar variation in f^0_E at Canberra. Quantity plotted is $200P_2/\text{mean}\,f^0_E$, 1937 to 1946.

$$200P_2/\text{mean}\,f^0_E = 0{\cdot}26\,\% \quad \text{and} \quad t_2 = 3{\cdot}8\,\text{hr}.$$

TABLE 2

season	$200P_2/\text{mean}\,f^0_E$	t_2 (lunar hours)
N	0·18	6·6
E	0·42	3·8
S	0·40	3·2

It will be observed that the variations of h'_E and f^0_E are approximately in phase, maximum heights occurring at the same time as maximum densities. In each quantity the maximum occurs about 2 to 3 hr. earlier in summer than in winter, the equinoctial maxima occurring at intermediate times.

2·2. *Region F_1*

The procedure followed for this region was identical with that described above for region *E*. For h'_{F_1} 7 years of data were used, covering the period 1938 to 1944. The average harmonic coefficients for each group of 4 months are displayed on the dial in figure 3. The grouping of the points shows the existence of a real 12-hourly lunar variation with maximum phase about 6 hr. after transit. The average coefficients for each season are shown in table 3. It is observed here from comparison with tables 1 and 2 that the seasonal run of t_2 appears anomalous for the *N* months, which might have been anticipated to yield a value of t_2 about 8 or 9 hr., and not 5·3 as found. The average coefficients for all seasons are $P_2 = 0{\cdot}83$ km., $t_2 = 6{\cdot}1$ lunar hours after transit.

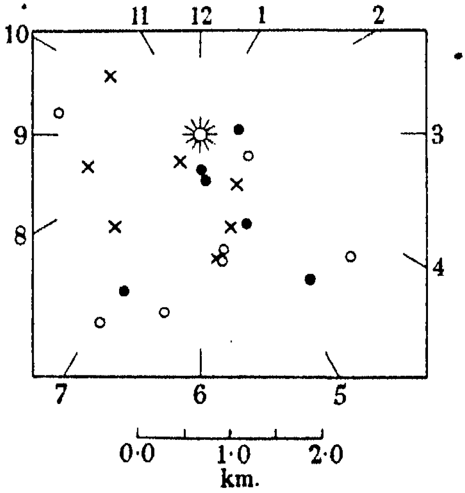


FIGURE 3. Harmonic dial for semi-diurnal lunar variation in h'_{F_1} at Canberra, 1938 to 1944.

TABLE 3

season	P_2 (km.) for h'_{F_1}	t_2 (lunar hours)
N	0·85	5·3
E	0·64	7·0
S	1·13	6·1

To determine the lunar variation in $f^0_{F_1}$, 10 years' data covering the period 1937 to 1946 were used. The results are displayed in figure 4 on the harmonic dial in the same manner as for f^0_E in figure 2. The average harmonic coefficients for each season are shown in table 4. The mean values of the harmonic coefficients for all seasons and all years are: for $200P_2/\text{mean}\, f^0_{F_1}$, 0·26 %; and for t_2, 7·5 hr.

It will be observed from tables 3 and 4 that while the lunar variations in h'_{F_1} and $f^0_{F_1}$ are approximately in phase for the S and E months, they are more nearly out of phase for the N months. The variation in $f^0_{F_1}$ for the latter months cannot be considered as satisfactorily determined, in view of the considerable scatter of the N points in figure 4, so that it would probably be unwise to attach significance to this result until further confirmed.

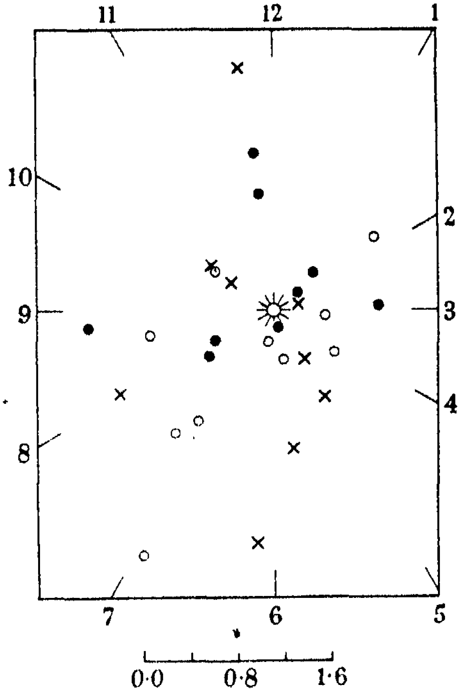


FIGURE 4. Harmonic dial for semi-diurnal lunar variation in $f^0_{F_1}$ at Canberra, 1937 to 1946. Quantity plotted is $200P_2/\text{mean } f^0_{F_1}$.

TABLE 4

season	$200P_2/\text{mean } f^0_{F_1}$	t_2 (lunar hours)
N	0·28	11·0
E	0·32	7·3
S	0·48	6·8

2·3. *Region* F_2

For the study of the lunar variations in h'_{F_2}, hourly values for the 4 years 1941 to 1944 were used. The results for the semi-diurnal harmonic are shown on the harmonic dial in figure 5. The mean seasonal coefficients are given in table 5.

TABLE 5

season	P_2 (km.) for h'_{F_2}	t_2 (lunar hours)
N	0·3	7·1
E	1·9	6·2
S	2·8	5·2

The mean values of the coefficients for all seasons and all years are $P_2 = 1{\cdot}60$ km. and $t_2 = 5{\cdot}7$ hr. The seasonal swing of t_2 already noted in the case of the E and F_1 regions also occurs in h'_{F_2}, winter maximum heights occurring about 2 hr. later than the summer maxima. Once again the summer value of P_2 is greater than the equinoctial, and the latter is greater than the winter P_2. In this case the winter P_2 is notably small, but it will be observed from figure 5 that this occurs because of the fact that two of the winter points on the dial are almost out of phase with the mean coefficient.

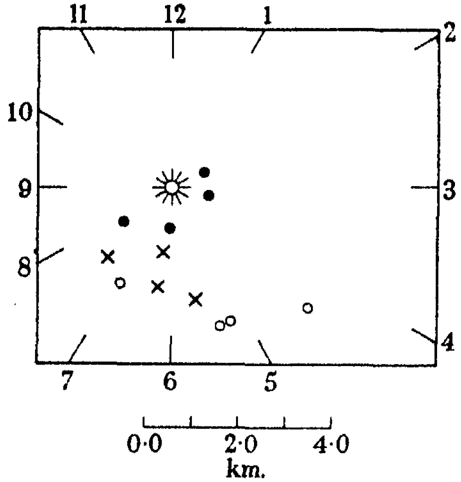


FIGURE 5. Harmonic dial for semi-diurnal lunar variation in h'_{F_2} at Canberra, 1941 to 1944.

In general, for the F_2 region, the minimum equivalent height h'_{F_2} is not a very reliable measure of the 'under-surface' of the region. This happens because the measurement obtained in practice is often vitiated by the presence of the F_1 region immediately below, which prevents the true scaling of the minimum height of the upper region. A better measure is the quantity $h^{\max.}_{F_2}$, which is the reading of h' at a frequency 0·834 time the critical penetration frequency of the region. According to Booker & Seaton (1939) $h^{\max.}$ obtained in this manner is a good measure of the height at which N(max.) occurs. In any event it may safely be assumed that lunar variations obtained from $h^{\max.}_{F_2}$ data refer to a somewhat higher level in the F_2 region than those similarly obtained from h'_{F_2} data.

For this study 3 years' data were used (1942 to 1944), the results being shown in figure 6 and table 6 in the usual way.

TABLE 6

season	P_2 (km.) for $h^{\max.}_{F_2}$	t_2 (lunar hours)
N	0·4	6·3
E	1·9	5·9
S	2·8	5·3

The mean variation over the whole period is shown graphically in figure 7. The mean coefficients for all seasons and all years are $P_2 = 1{\cdot}68$ km., $t_2 = 5{\cdot}7$ hr. Clearly, the results obtained for $h^{\max.}_{F_2}$ agree very closely with those for h'_{F_2}, and it is hardly possible to attach significance to the small differences found. For example, when the points for 1941 are omitted from the h'_{F_2} dial, in order to make the results strictly comparable with those for $h^{\max.}_{F_2}$, the mean t_2 for N months in h'_{F_2} becomes 6·0. (It was for this reason, as the analysis progressed, that it was not thought worth while

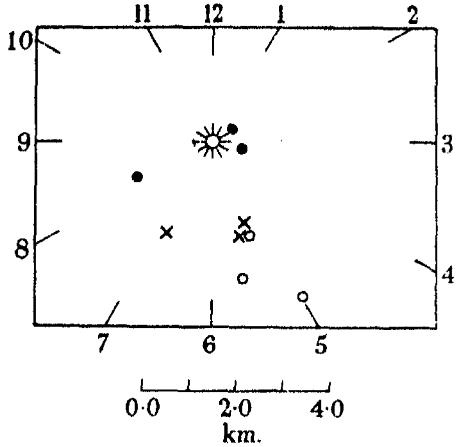


FIGURE 6. Harmonic dial for semi-diurnal lunar variation in $h^{\max.}_{F_2}$ at Canberra, 1942 to 1944.

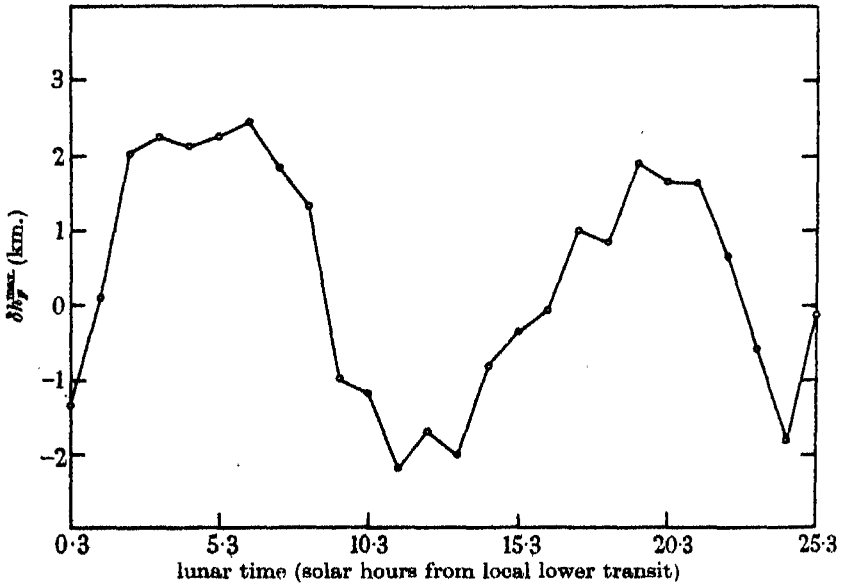


FIGURE 7. Mean lunar variation in $h^{\max.}_{F_2}$ at Canberra, 1942 to 1944.

to analyze 1941 for $h^{\max.}_{F_2}$, a year whose data had already been used in deriving the h'_{F_2} variation.) It is to be recalled, however, that a significant difference between the h'_{F_2} and $h^{\max.}_{F_2}$ lunar variations was found in II for Huancayo, a site at which both these quantities were found to be considerably greater than at Canberra, and where the total ranges of solar variation of h'_{F_2} and $h^{\max.}_{F_2}$ are considerably greater than at Canberra.

The lunar variation of $f^0_{F_2}$, the critical penetration frequency of the region, was studied in the hourly values of that quantity for the three years 1942 to 1944. The results (for $200P_2$/mean $f^0_{F_2}$) are shown in the harmonic dial in figure 8. The mean coefficients are shown in table 7.

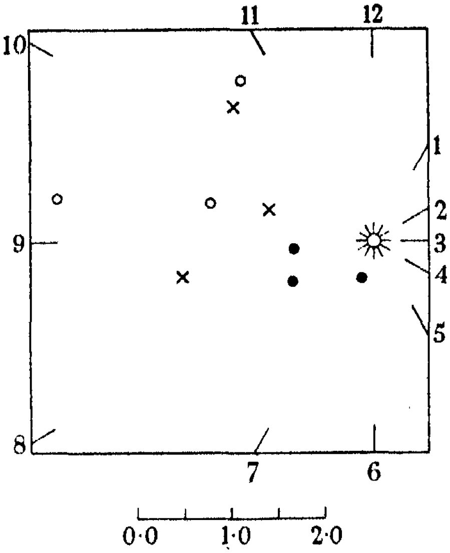


FIGURE 8. Harmonic dial for semi-diurnal lunar variation in $f^0_{F_2}$ at Canberra, 1942 to 1944. Quantity plotted is $200P_2$/mean $f^0_{F_2}$.

TABLE 7

season	$200P_2$/mean $f^0_{F_2}$	t_2 (lunar hours)
N	0·7	8·1
E	1·6	9·5
S	2·4	9·6

The mean lunar variation of $f^0_{F_2}$ over the whole period is shown graphically in figure 9. The mean coefficients for all seasons and all years are $P_2 = 1{\cdot}5\,\%$ and $t_2 = 9{\cdot}4$ hr. As in table 3, the value of t_2 for the *N* months seems anomalous: instead of 8·1 hr. a value of 10 to 11 hr. might have been anticipated from the run of t_2's in the other tables.

In II it was found that the harmonic coefficients of the lunar variations in the F_2 region depended markedly on solar time. A similar analysis has been made of the Canberra data for $h^{\max.}_{F_2}$ and $f^0_{F_2}$. For this purpose the lunar harmonic coefficients were derived for each separate hour of the solar day. The results for $h^{\max.}_{F_2}$ are shown graphically in figure 10 and for $f^0_{F_2}$ in figure 11. In each figure the variations of P_2 and t_2 with solar time are shown separately. Since each point on these figures has been derived from the harmonic analysis of only 37 lunar days, it might be anticipated that they would show considerable statistical scatter. Actually the scatter is surprisingly small, particularly in the amplitude P_2. As might be anticipated, the t_2's show considerable scatter at solar times when P_2 is small, but at other times they too show a smooth variation. In the afternoon and evening hours the t_2's are very

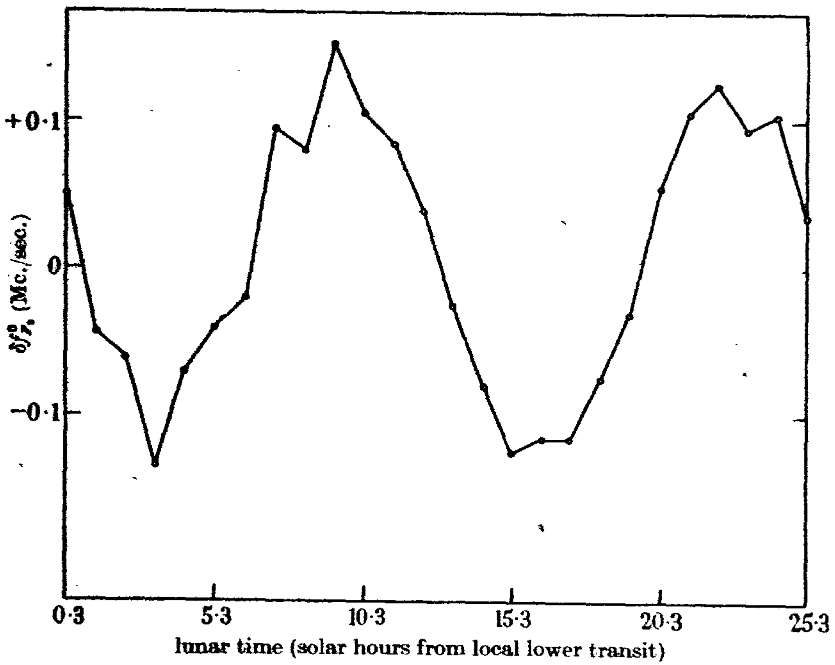


FIGURE 9. Mean lunar variation in $f^0_{F_2}$ at Canberra, 1942 to 1944.

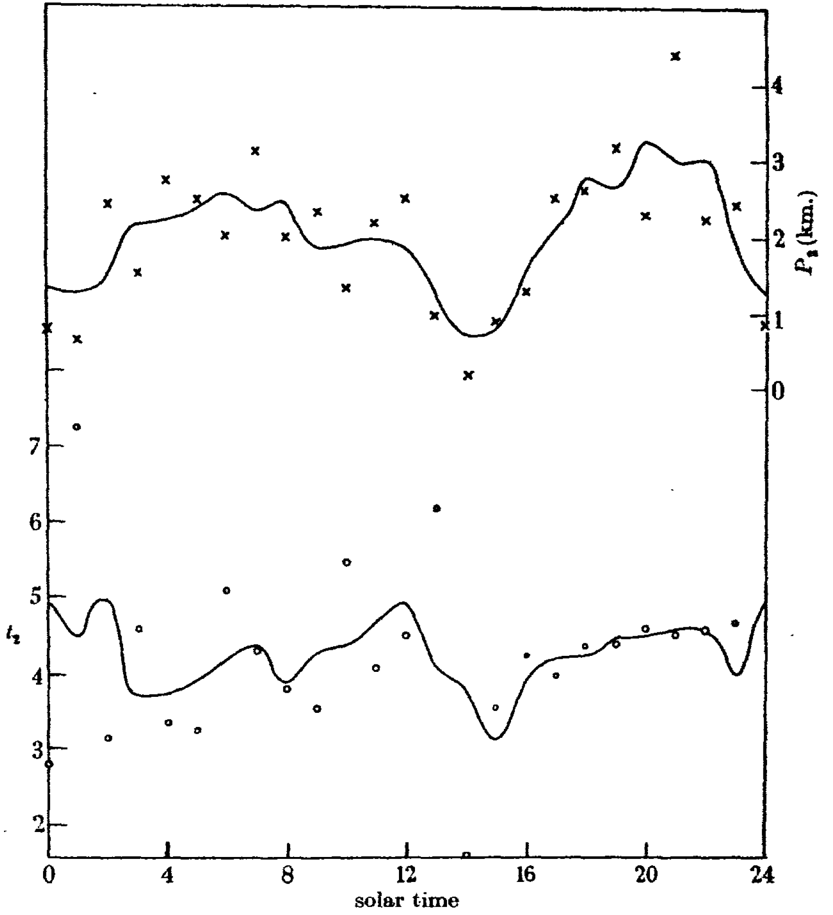


FIGURE 10. Variation of P_2 (upper curve) and t_2 (lower curve) with solar time for $h^{\max.}_{F_2}$ at Canberra, 1942 to 1944.

steady, while for $f^0_{F_2}$ they show a smooth rise from 01 to 08 solar hours. (The latter variation is very similar to that found in II for the corresponding quantity at Huancayo.)

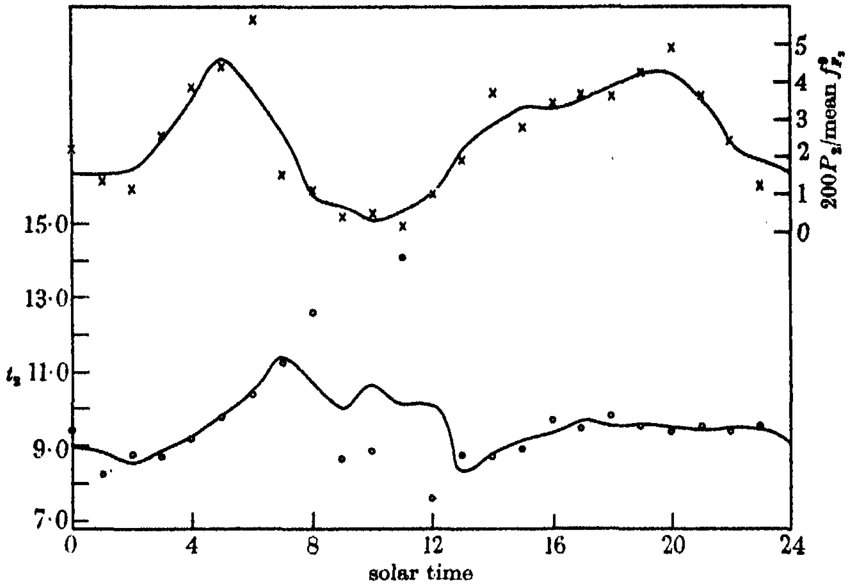


FIGURE 11. Variation of P_2 (upper curve) and t_2 (lower curve) for $f^0_{F_2}$ at Canberra, 1942 to 1944.

In each figure a smooth curve has been drawn through the centres of gravity of each successive group of three points, thus partially eliminating the scatter. It seems clear that, apart from the smooth variation in t_2 (figure 11) at 01 to 08 solar hours mentioned above, there is no clearly marked variation in t_2 with solar time for either $h^{\text{max.}}_{F_2}$ or $f^0_{F_2}$. On the other hand the amplitudes P_2 in both these quantities (upper curves, figures 10 and 11) appear to show a clearly marked semi-diurnal variation in solar time. In view of its probable theoretical importance this variation has been harmonically analyzed with results as follows:

(1) $h^{\text{max.}}_{F_2}$: $P_2 = 0{\cdot}84$ km., $t_2 = 7{\cdot}6$ solar hours.

(2) $f^0_{F_2}$: $P_2 = 1{\cdot}44\,\%$, $t_2 = 5{\cdot}3$ solar hours.

3. DISCUSSION OF RESULTS

The results described in the previous section show that there are observable lunar semi-diurnal variations in all the main regions of the ionosphere. These occur not only in the heights of these regions, but also in the maximum electron densities. The latter fact alone would appear to rule out any simple explanation of the phenomena in terms of simple tidal rising and falling of the isobaric surfaces in the upper atmosphere. While such an explanation might account for the lunar *height* variations of a 'Chapman' region, the maximum density of which occurs at a definite level of atmospheric pressure, it would not account for variations in maximum electron density at that level.

In general it will be observed that while for the E and F_1 regions the lunar variations in h and N_m (N(max.)) are approximately in phase, largest values of N_m occurring when the region is highest, for the F_2 region the variations in h and N_m are more nearly in phase quadrature, maximum of N_m occurring 3·7 lunar hours after the maximum in h.

In I and II a beginning was made to the theory of the vertical distortion of a Chapman region by vertical ionic drift. The fundamental equation of this theory is

$$\frac{\partial N}{\partial t} = I(z,t) - \alpha N^2 + \frac{\partial (Nv)}{\partial z}, \tag{1}$$

where $I(z,t)$ is the rate of ion production per c.c., α is the effective recombination coefficient, z is the 'reduced' height, and v is the (downwards) vertical ionic (and electronic) velocity, which is periodic in time, and which may vary in both amplitude and phase with z. Or, say, $v = v_0(z) \sin(\omega t + \sigma z)$.

In the previous papers the theory has been applied to the F_2 region, and it has been considered justifiable, as a first approximation, to neglect the recombination term. Also, night-time conditions only have so far been considered, so that equation (1) has been reduced to the relatively simple equation of continuity

$$\frac{\partial N}{\partial t} = \frac{\partial (Nv)}{\partial z}. \tag{2}$$

It has been shown that, for conditions where v varies only in amplitude with z, an ionized region undergoes height and density variations which are either in or out of phase, maximum height occurring simultaneously with either maximum or minimum density. On the other hand, when v changes phase with z, the variations in height and density are approximately in phase quadrature. In the latter case, if v varies harmonically and semi-diurnally, the height of the region is greatest approximately 3 hr. after the maximum upward velocity occurs at the level of N_m, i.e. when the vertical velocity at the N_m level is zero. Physically, it is clear why this should happen. The region continues to rise while there is an upward velocity of ions, and reaches maximum height about the time when the ions are just starting to move downwards. It is also found theoretically (II) that N_m reaches its maximum value either when v has its maximum (downward) value (σ +ve), or when v has its maximum upward value (σ −ve). It follows that the variation in z_m (the height at which N_m occurs) leads that of N_m by about 3 hr. when σ is positive, and lags 3 hr. when σ is negative. Symbolically, it was shown in II that when only the phase of v varies with z

$$N_m = 1 + \frac{\sigma v_0}{\omega} \sin(\omega t + \sigma z_m)$$

and $$z_m = \frac{v_0}{\omega} \cos(\omega t + \sigma z_m), \quad \text{provided } \sigma v_0 \ll \omega.$$

The results described in §2 above show that, for the F_2 region, the N_m and z_m variations are substantially in phase quadrature, so that the above theory can probably be safely applied here.

On the other hand, the N_m and z_m variations in the E and F_1 regions are found above to be substantially in phase agreement, and it seems clear that the lunar tidal movements and distortions of these regions are not conditioned by the vertical phase gradient of v. This conclusion is hardly surprising, since the effective recombination coefficient α is considerably greater in the E and F_1 regions than it is in F_2, and can no longer be ignored in equation (1). It was shown in I that the ratio of the last term in equation (1) to the recombination term is of order $v/H\alpha N$, where H is the 'height of the homogeneous atmosphere' at the level concerned. In the F_2 region v appears to be of order 100 cm./sec., H about 5×10^6 cm., α about 10^{-11} or even 10^{-12}, and N about 10^6/cm.^{3}. Hence $v/H\alpha N > 1$ and the tidal term in the equation is particularly important. On the other hand, for the E region, H is say 10^6 cm., α is 10^{-8} (Best, Farmer & Ratcliffe 1938) and N about 10^5, so that $v/H\alpha N$ is less, and possibly much less, than unity. The ionizing term $I(z, t)$, and the recombination term, may therefore be expected to dominate the equation, and considerations of continuity will be secondary. A discussion of equation (1) when all terms are included will be given elsewhere, but for present purposes it may safely be assumed, on physical grounds, that the height variations in regions of this type will be approximately in phase with v, possible lagging 1 to 2 hr. for intermediate values of α.

In general, the results of the theory of variations of h_m and N_m developed in I and II may be summarized briefly in the following way, which provides simple rules for examining the meaning of a given variation.

Rule (1). An ionized region moves in the direction of the ionic drift, upwards when that is upwards, and downwards when it is downwards. Exceptions to this rule occur only when the vertical gradient of v (in either amplitude or phase) is unusually great, or when α (and I_0) are particularly large. In the latter case the velocity of movement of the region may lead v by 2 to 3 hr.

Rule (2). Variations in N_m are proportional to the height gradient of v (in either amplitude or phase), and their sign (+ or −) depends on the sign of this gradient.

3·1. *Discussion of E region results*

The most notable feature of the results for this region is the phase of the variation in h'_E (5·1 hr.), which is almost exactly opposite to that obtained by Appleton & Weekes (1939) for the same region in England (11·25 hr.). The amplitude of the variation at Canberra (0·19 km.) is considerably smaller than that obtained in England (0·85 km.), but figure 1 shows that there can be no doubt that the phase at Canberra is in opposition to that in England. (This result is confirmed by the analysis of data from Brisbane (lat. 27·5° S.) which gives $P_2 = 0{\cdot}5$ km., and $t_2 = 4{\cdot}5$ lunar hours for the E region.)

This result is of considerable importance theoretically. According to any simple theory based on the tidal rise and fall of isobaric surfaces, the phase of the variation should be the same all over the globe, when expressed in hours from local transit. On the other hand, it is a necessary consequence of the theory developed in I, which attributes region movements to the 'dynamo' electric forces in the atmosphere, that the phase of this variation should change by π in the vicinity of latitude 35°. The present result appears therefore to give strong support to the theory and suggests that Canberra (lat. 35·3° S.) is on the equatorial side of this 'node'. The smallness of the variation at Canberra is now readily understood, since this site cannot be very far north of the node (cf. figure 5, curve II in I). The larger value of P_2 at Brisbane, mentioned above, is also now readily understood.

Until equation (1) has been fully solved for the E region it would be unwise to draw any deductions from the small but definite lunar variation in f^0_E. (In this respect the height gradient of α may also have to be considered.)

It is possible, however, to make one safe deduction from the evidence as a whole. There is little doubt that the maximum upward velocity of ions in the E region occurs, on the equatorial side of the node, due to lunar influence, between 2 and 5 hr. after lunar transit. (The extension to 2 hr. is made in order to allow for the uncertainty mentioned at the end of Rule (1) in the previous section.) Now this maximum upward velocity occurs when the 'dynamo' electric force in the E region has its maximum value *Eastwards*. It is possible then to compare the phase of the electric forces thus found in the E region from ionospheric evidence, with that which occurs in the (unknown) region of the upper atmosphere which is mainly responsible for the lunar magnetic variations. The overhead system of currents which is responsible for these magnetic variations is now well known, and is depicted in the form of current sheets by Chapman & Bartels (1940, vol. 1, p. 262). A study of these sheets shows that, on the equatorial side of the current focus, which corresponds to our node, the currents have a maximum *Westerly* component 3·5 hr. after lunar transit. It is seen therefore that the electric forces found in the E region are almost exactly out of phase with the currents producing the lunar magnetic variations. It follows that the lunar magnetic variations are not produced mainly in the E region.

3·2. *Discussion of F_1 region results*

The results for this region, taken as a whole, are generally similar to those for region E. The most noticeable differences are the increased amplitude of the variation in h', a small but definite retardation in the phase of this variation, and a marked retardation in the phase of the f^0 variation. In all these respects the F_1 region shows characteristics intermediate between those for the E and F_2 regions.

In II the lunar variation in $h^{\text{max.}}_{F_1}$ at Huancayo, near the magnetic equator, was determined from three years' data (1942 to 1944). P_2 was there found to be 0·5 km., and t_2 4·0 lunar hours. These results are not strictly comparable with those for Canberra, which refer to h'_{F_1} for the years 1938 to 1944. When the years 1938 to 1941 are eliminated from the Canberra data the result is $P_2(h'_{F_1}) = 1 \cdot 1$ km. and $t_2 = 5 \cdot 5$ hr. In general the level of $h^{\text{max.}}_{F_1}$ is some 15 km. higher than that of h'_{F_1}. The results obtained at the two sites are therefore consistent with the interpretation that: (*a*) both sites are on the equatorial side of the node, the phases of the F_1 variations being substantially the same at both places; (*b*) the phase of v is advancing at the F_1 level, that of $h^{\text{max.}}_{F_1}$ being approximately 1 hr. ahead of that of h'_{F_1}.

From reasoning identical with that used at the end of the previous section it is concluded that the lunar magnetic variations are not produced mainly in the F_1 region. As in the case of the E region, the evidence shows that the currents flowing in the F_1 region are almost exactly out of phase with those required to produce the observed magnetic variations.

3·3. *Discussion of F_2 region results*

The first points to be noticed for this region are the increased amplitude of the variations, as compared with the lower regions, and the fact that the variations in h and f^0 are now in approximate phase quadrature. The latter fact is interpreted as showing that the variations in f^0 are governed mainly by the vertical phase gradient in v. On this interpretation, the theory given in II shows that the phase gradient σ is $+2P_2(f_0)/\text{mean}\, f_0 . P_2(h^{\text{max.}})$. Substituting the mean values of these harmonic coefficients from tables 6 and 7, the magnitude of σ is determined for each season. At the same time the values of v_0 are determined from the relationship

$$P_2(h^{\text{max.}}) = v_0/\omega,$$

where ω is 0·5 radian/hr. The results of these calculations are set out in table 8. It is doubtful if the apparent sharp increase in v_0 with height is real. So sharp a gradient might be expected to govern the variation in N_m, and to bring it into (or out of) phase with the h variation. It seems more likely that the apparent increase in v with height is due to the rapid decrease in importance of the recombination and ionizing terms in equation (1), and the consequent removal of their restraining influence on bodily movement of the F_2 region. On the other hand the apparent decrease in σ with height is probably real, since it is derived from the ratio $P_2(f_0)/P_2(h^{\text{max.}})$, and each of these coefficients will be influenced to roughly the same extent by any change in importance of the ionizing and recombination terms.

TABLE 8

season	v_0 (km./hr.)	σ (lunar hours/km.)	mean $h^{\text{max.}}_{F_2}$
N	0·2	+0·035	298
E	0·9	+0·017	314
S	1·4	+0·017	333

Another reason for caution in interpreting the large seasonal change in v_0 is as follows. There is reason to believe (§ 3·1) that Canberra is situated near to the current focus of the world-wide system of ionospheric currents, i.e. near to the node in v_0. There is good reason to believe that the position of this node shifts north and south seasonally by a few degrees of latitude. The small N month variations in the F_2 region may simply mean that in these months the node is almost overhead.* Once again, however, the calculated values of σ will be unaffected, since this seasonal shift will affect the variations in f^0 and $h^{\text{max.}}$ equally.

The lack of significant difference in the t_2's for the $h^{\text{max.}}_{F_2}$ and h'_{F_2} variations is not surprising. According to the values of σ in table 8 the expected difference is about 0·5 hr., which is just about twice the probable error in the determination of each t_2.

The relationship of $t_2(h'_{F_1})$ to $t_2(h'_{F_2})$ is interesting. According to the values found for σ it might at first sight be anticipated that the latter would be about 100 × 0·035 or 3·5 hr. earlier than the former. In fact, it is only 0·4 hr. earlier. It must be remembered, however, that a retardation of $\frac{1}{2}\pi$ occurs in the h variation in passing upwards to a region where the recombination and ionizing terms no longer mainly control position. Similar considerations must apply to the relations between $t_2(f^0_{F_1})$ and $t_2(f^0_{F_2})$. The latter value should be obtained from the former by advancing it 3·5 hr. (to 4·0 hr.) and then changing phase by π (to 10·0 hr.).*

---

* Strong support for this interpretation is provided by the $f^0_{F_2}$ lunar variations at Watheroo, Western Australia, which have been found to change phase by π from summer to winter.

The maximum *upward* velocity of ions occurs 3 hr. before the maximum height, i.e. at 2·7 lunar hours. The evidence of the lunar magnetic variations shows that ions must move *downwards* with maximum velocity at this hour in the region which mainly produces this variation. It is concluded that the lunar magnetic variation is not produced mainly in the F_2 region.

A comparison of the present results for the F_2 region has been made with those described in II for Huancayo. In view of the strong dependence of P_2 and t_2 on solar time at the latter site, it does not appear possible to draw any useful conclusions. $P_2(f^0_{F_2}, h^{\text{max.}}_{F_2})$ Huancayo are in general much larger than the corresponding quantities for Canberra, and are so variable with solar time that it would be unwise to select their mean values over the solar day for comparison with other sites.

3·4. *Discussion of the luni-solar variations in the F region*

The variations of P_2 and t_2 with solar time, as exhibited in figures 10 and 11 may be briefly summarized by the statement that there are no clearly marked variations in $t_2(f^0_{F_2}, h^{\text{max.}}_{F_2})$, and well marked semi-diurnal variations in $P_2(f^0_{F_2}, h^{\text{max.}}_{F_2})$. A non-linear theory of these variations was developed in II for cases where v_1, the solar ionic drift velocity, was much greater than v_2, the corresponding lunar velocity. In the case of the corresponding Canberra variations, the degree of non-linearity present is obviously smaller, and a simple understanding of the results might be obtained by expressing the variations in a power series. Thus

$$N_m = 1 + a\frac{\sigma v_0}{\omega}\sin\omega t + b\frac{\sigma v_0^2}{\omega^2}\sin^2\omega t + \dots \tag{3}$$

and

$$z_m = \frac{a' v_0}{\omega}\cos\omega t + b'\frac{v_0^2}{\omega^2}\cos^2\omega t + \dots. \tag{4}$$

If the total drift be due to the combined action of the solar and lunar drifts, whose parameters are denoted by the subscripts 1, 2 respectively, then

$$N_m = 1 + a\frac{\sigma_2 v_2}{\omega_2}\left(1 + 2\frac{b}{a}\frac{\sigma_1 v_1}{\omega_1}\sin\omega_1 t\right)\sin(\omega_2 t + \theta) \tag{5}$$

and

$$z_m = a'\frac{v_2}{\omega_2}\left(1 + 2\frac{b'}{a'}\frac{v_1}{\omega_1}\cos\omega_1 t\right)\cos(\omega_2 t + \theta), \tag{6}$$

where θ is an unknown phase constant. In deriving these expressions only terms of frequency ω_2 have been selected, since it is desired to compare the theory with the results of an analysis of data which selects only these terms.

---

* This latter change is made up of a retardation of $\frac{1}{2}\pi$ as for the h variation, plus a further retardation of $\frac{1}{2}\pi$ expressing the change from the F_1 region, where the f^0 variation is conditioned by the height gradient of either α or of *amplitude* in v, to the F_2 region, where this variation depends on the *phase* gradient in v.

The above analysis shows that there should appear solar semi-diurnal variations in the amplitudes of the lunar variations in f^0 and $h^{\text{max.}}$. Moreover, these solar variations should be in approximate phase quadrature. (It was found above, at the end of § 2·3, that these solar variations differ in phase by 2·3 hr.) At first sight therefore, a satisfactory interpretation of the luni-solar variation is obtained from the above simple theory. There is little doubt, however, that this gives only a partial representation of the facts. Difficulties appear when attempts are made to align the phases of these solar amplitude variations with the fundamental solar variation. It is probable that the variations of v_2 and σ_2 with height are important in this connexion. The solar drift almost certainly moves the F_2 region through levels where v_2 and σ_2 are changing markedly. For example, it appears from the evidence in II that in the F_2 region above Huancayo σ_2 is negative in sign in daylight hours. At this time the F_2 region is particularly high (*c.* 400 km.). Such high values of $h^{\text{max.}}_{F_2}$ are never attained at Canberra and σ_2 appears to be always positive. Nevertheless the evidence suggests that σ_2 becomes zero in the neighbourhood of 350 km. The small values of $P_2(f^0_{F_2})$ exhibited in figure 11 at noon and midnight might therefore equally well be attributed to this gradient in σ_2. A decision on this point will require the support of other evidence.

This work has been carried out as part of the research programme of the Council for Scientific and Industrial Research. It has been done while the author was visiting the Commonwealth Observatory, Mount Stromlo. The thanks of the author are due to Dr R. Woolley for the provision of facilities and for useful discussion. The Radio Research Board has provided a team of computers led by Misses A. McGillivray and H. Lee, whose assiduity under the direction of Miss B. Hardwick deserves special thanks. Mr W. L. Price has carefully checked and criticized all the theoretical work in this and the two previous papers.

REFERENCES

Appleton, E. V. & Weekes, K. 1939 *Proc. Roy. Soc.* A, **171**, 171.
Best, J. E., Farmer, F. T. & Ratcliffe, J. A. 1938 *Proc. Roy. Soc.* A, **164**, 96.
Booker & Seaton 1939 *Trans. Wash. Meet. Ass. Terr. Mag. and Elect.* p. 503.
Chapman, S. & Bartels, J. 1940 *Geomagnetism.* Oxford: Clarendon Press.
Martyn, D. F. 1947*a* *Proc. Roy. Soc.* A, **189**, 241.
Martyn, D. F. 1947*b* *Proc. Roy. Soc.* A, **190**, 273.

Atmospheric tides in the ionosphere

IV. Studies of the solar tide, and the location of the regions producing the diurnal magnetic variations

By D. F. Martyn, D.Sc.

Radio Research Board, Australian Council for Scientific and Industrial Research

(*Communicated by Sir Edward Appleton, F.R.S.—Received* 3 *April* 1948)

Solar tidal effects are sought in ionospheric data from representative latitudes. They are *found in the E, F_1 and F_2 regions. For the F_2 region the seasonal semi-diurnal harmonics* arising from P_3^2 are comparable with those from P_2^2. The amplitudes and phases of these *harmonics are determined, and they are used to interpret the global distribution of F_2* ionization. Evidence is found that the lunar magnetic variation is produced mainly by *currents in the D region of the ionosphere, but is opposed by the corresponding currents in* the E and F_1 regions. This finding is submitted to an independent test, using McNish's *evidence of the effects of solar flares on the magnetic variation.*

1. Introduction

In three previous papers some consequences of solar and lunar tidal motions in the ionosphere have been investigated. In the first of these (I, Martyn 1947) a theory was propounded which attributed the major anomalies in the behaviour of the F_2 region to the action of the solar atmospheric tide. The theory showed that this tide, when account was taken of the presence of the earth's magnetic field, could cause notable vertical distortion of an ionized region. It is not easy to separate the effects of the solar tide from the solar ionizing influence, since both have fundamentally the same periodicity. Accordingly two papers (II, Martyn 1947; III, Martyn 1948) were devoted to the study of lunar variations in the ionosphere. This study has appeared to verify the main features of the theory of region distortion developed in I. The present paper resumes the study of the solar tide. In particular, an attempt is made to determine the phase of this tide in the main regions of the ionosphere, a determination which has already been achieved in III for the lunar tide. Success in this attempt would permit conclusions regarding the seats of the currents which produce the solar and lunar diurnal magnetic variations.

2. The distortion of an equilibrium 'Chapman region' by vertical ionic drift

It is generally assumed that the diurnal and seasonal variations of the maximum electron densities (N_m), and the heights at which these occur ($h^{\text{max.}}$), are adequately described, for the E and F_1 regions, by Chapman's theory (1931). It will appear below that there are small but significant departures from 'Chapman region' behaviour in both these regions. These anomalies become evident when hourly data from ionospheric observatories are averaged over periods of the order of 10 years. Reasons will be advanced why these anomalies are attributed to the vertical ionic (and electronic) drift produced by the 'dynamo' electric forces associated with the solar atmospheric tide.

In I and II the theory of the distortion of a Chapman region by such vertical drift was worked out, with the limitations that ion production and recombination were neglected. Under these restrictions the results could only be applied to the F_2 region during the hours of darkness. For the purposes of the present paper, in which the E and F_1 regions are also studied, it is clearly necessary to extend the theory to the case of a Chapman region in which ion production and ion recombination are in quasi-equilibrium. In fact, whereas in the F_2 region at night vertical ionic drift is found to be the most important cause of variations in $h^{\text{max.}}$ (and sometimes even in N_m), in the E and F_1 regions such drift can only be considered as a small perturbing term in Chapman's equation. The equation for N then becomes

$$\frac{\partial N}{\partial t} = I(z,t) - \alpha N^2 + \frac{\partial (Nv)}{\partial z}, \tag{1}$$

where $I(z,t)$ is the rate of ion production per cm.^{3}, α is the effective recombination coefficient, z is the 'reduced' height, i.e. the height measured in units of H the 'height of the homogeneous atmosphere', and v is the vertical ionic (and electronic) velocity,* measured positively downwards. v is periodic in time, and may vary with z in both amplitude and phase, i.e. $v = v_0(z)\sin(\omega t + \sigma z)$.

The solution of equation (1) for N_m and $h^{\text{max.}}$ is obviously difficult. For the purposes of the present paper it is desired to find the relationship between the phases of the vertical drift and the consequent small disturbances in N_m and $h^{\text{max.}}$, together with an approximate measure of their magnitude. A solution has been made by Mr C. B. Kirkpatrick, for the numerical values of the parameters shown in table 1, which are taken as sufficiently representative of E-region conditions at the equator at the equinoxes. (The full details of Mr Kirkpatrick's work will appear elsewhere.)

TABLE 1

α	σ	H	ω	v_0	N_m (noon)
10^{-8} cm.^{3}/sec.	4°	11 km.	30°/hr.	22·5 km./hr.	$1{\cdot}5 \times 10^5$

In figure 1 are shown the departures of N_m and $h^{\text{max.}}$ from the theoretical 'Chapman' values, occasioned by the last term in equation (1), for the set of parameters in table 1. The uppermost curve shows the percentage departure of N_m, the middle curve the assumed time variation in v, and the lowest curve the departure of $h^{\text{max.}}$ in km. It will be observed that the displacement of $h^{\text{max.}}$ follows closely the time variation of v, *and has the same sense as* v. This result, although it seemed likely on physical grounds, could hardly have been assumed without making a complete solution of equation (1). Considerable use will be made of it below. The departures of N_m are more complicated, and appear to depend, in both amplitude and phase, on the height gradient of v. This again is physically likely, and agrees with the conclusions of the former papers in this series.

---

* It is to be noted that the assumption that v is the same for ions and electrons (see I) is valid when $\omega^2/(\nu^2+\omega^2)$ is approximately unity for each particle, ω being the gyro-magnetic frequency and ν the collision frequency. In the F_2 region this condition is satisfied closely, in the F_1 region approximately. In the E region $\omega^2/(\nu^2+\omega^2)$ is considerably less than unity for ions. Consequently electrons in the E region will drift vertically at a faster rate than ions.

If the resulting current gave rise to polarization the electrons might be slowed down considerably.

It is more likely that a circuit will be completed in the E-W vertical plane, since upward drift at a given place is accompanied by downward drift 90° of longitude away. This current will produce no appreciable magnetic effects at the ground, but the partial polarization necessary to complete the circuit will reduce the electronic drift velocity appreciably.

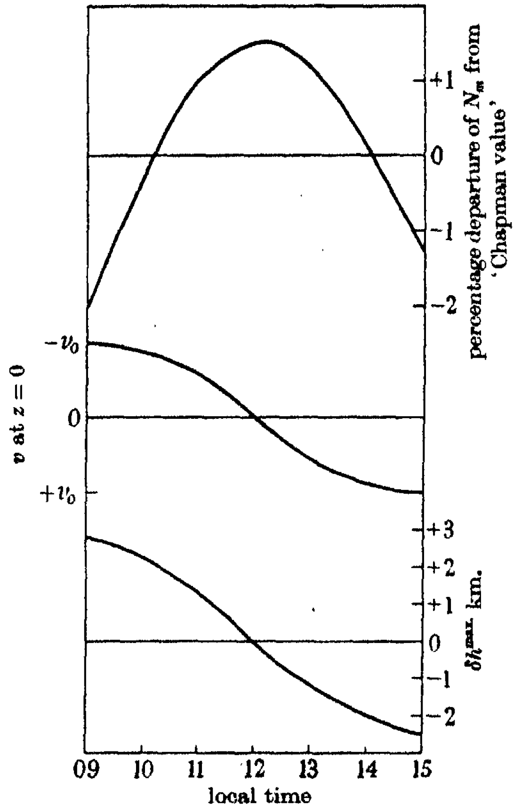


FIGURE 1. The theoretical departures of N_m (top curve) and $h^{\max.}$ (bottom curve) from 'Chapman values' produced by vertical tidal drift (middle curve).

3. THE F_1 REGION

The theory of the previous section indicates that, for likely values of solar tidal velocities in the ionosphere, there should be observable departures in the diurnal variations of $h^{\max.}$, for the E and F_1 regions, from those predicted by Chapman's theory. It is customary at ionospheric observatories to scale, not $h^{\max.}$, but h', the minimum equivalent height of these regions. For the F_1 region it is found that h' as so measured is, on the average, about 5 km. less than $h^{\max.}$, and shows no systematic difference from $h^{\max.}$ in its diurnal variation. It would appear, therefore, to be safe to take the normally scaled values of h' for the purpose of assessing the diurnal variations in $h^{\max.}$.

Figure 2a shows the average diurnal variations of $h'F_1$ at Canberra (lat. 35·3° S, long. 149° E) for the years 1937 to 1947 inclusive. In this figure the average curve for each season is shown separately. Curves marked N refer to the months around northern solstice (May, June, July, August). The curves for southern solstice are marked S, and those for the four equinoctial months E. The probable error of each measurement of $h'F_1$ is 5 km. Since each plotted point on the figures is the mean of approximately 1320 hourly readings, it follows that the probable error of each plotted point is 0·14 km. According to Chapman's theory $h^{\text{max.}}$ should vary closely as $\log_e(\sec\chi)$, where χ is the solar zenith angle. Hence the variation of $h^{\text{max.}}$ should be symmetrical about noon.

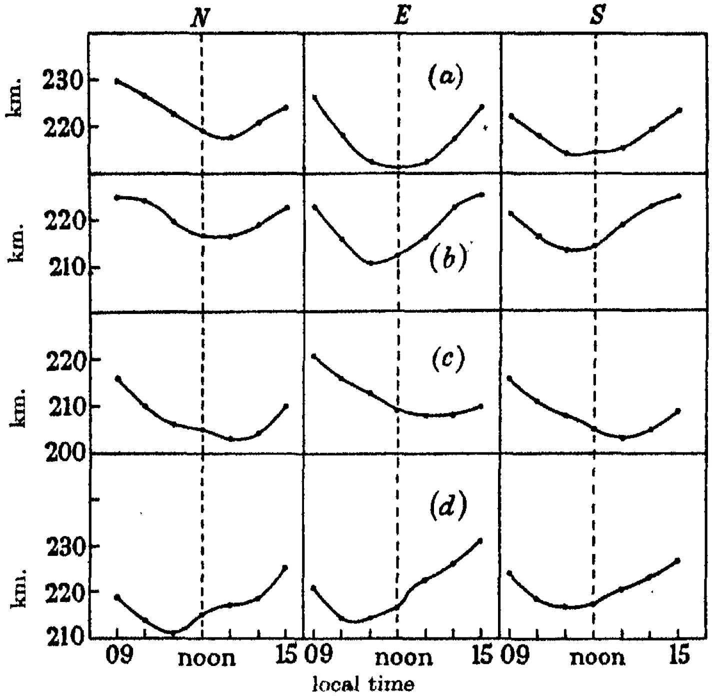


FIGURE 2. Average diurnal variations (09 to 15 hr.) of $h'F_1$ for each season (N, E, S) at (a) Canberra 1937 to 1947, (b) Watheroo 1938 to 1946, (c) Huancayo 1937 to 1947, (d) Washington 1934 to 1947.

It is apparent that there are notable departures in the $h'F_1$ diurnal variation at Canberra from that predicted by Chapman's theory. In the winter months $h'F_1$ is anomalously high in the morning (or low in the afternoon). This anomaly is roughly reversed in the summer months. The fact that these anomalies are unsymmetrical about noon would appear to rule out any possibility of their being of instrumental origin. For example, there is always some uncertainty in scaling $h'F_1$ because of the variable length of the horizontal portion of the (h'-frequency) curve. This length is determined by the difference between f^0F_1 and f^0E. Since this latter difference is almost perfectly symmetrical about noon, any error in scaling $h'F_1$ should also be symmetrical about noon.

A further test of the reality of this anomaly in $h'F_1$ has been made by plotting the mean diurnal variations of $h'F_1$ at Watheroo (lat. 30·3° S, long. 115·9° E) which has almost the same magnetic latitude as Canberra. This observatory operates ionosphere sounding equipment of a different pattern from that in use at Canberra, and the records there are of course scaled by different personnel. The results of this plot for the years 1938 to 1946 are shown in figure 2*b*. It will be seen that the pattern of the anomalies in $h'F_1$ at Watheroo is closely similar to that at Canberra.

If the anomalies in $h'F_1$ are due to the solar tide, then it would appear that at Canberra and Watheroo the drift velocity due to this tide is upwards in winter in the late morning, and downwards at the same time in summer. Now it is known that the focus of the ionosphere current system which produces the solar magnetic variations passes almost overhead in these latitudes. It is also known that this focus moves in latitude seasonally by several degrees, so that in N months an observatory in these latitudes can lie on the equatorial side of the focus, while in S months it lies on the polar side. According to the theory of vertical drift (I), the phase of this drift on the equatorial side of the focus is different by 6 hr. from that on the polar side. (This point has already been established in III, from observational evidence, for the lunar drift.) If this interpretation of the seasonal swing of the $h'F_1$ anomaly at Canberra and Watheroo be correct, then it should follow that observatories at low latitudes should show a downward drift in the F_1 region in the afternoon at all seasons. Conversely high-latitude observatories should show a downward drift in the F_1 region in the morning at all seasons.

The most suitable data for testing this point would appear to be those obtained at Huancayo (lat. 12·0° S, long. 75·3° W) and Washington (lat. 38·6° N, long. 77·5° W). Both these observatories have been operating for over 10 years. The average diurnal variations of $h'F_1$ at Huancayo for the years 1937 to 1947 are shown for each season in figure 2*c*. It will be observed that the F_1 region is anomalously high in the morning for all seasons (or low in the afternoon).

The corresponding $h'F_1$ variations at Washington for the years 1934 to 1947 are exhibited in figure 2*d*. It will be observed that the F_1 region is raised in the afternoon and depressed in the morning at all seasons. Similar results have been obtained from smaller series of observations at other sites (e.g. Brisbane (equatorial) and Ottawa (polar)).

This would appear to be strong evidence in favour of the interpretation of the $h'F_1$ anomalies as due to vertical drift associated with the 'dynamo' electric forces produced by the solar tide. If this interpretation be correct, then since maximum upward drift is associated with maximum eastward electric force in the ionosphere, and so (very closely) with the maximum value of horizontal magnetic force at the ground, it is easy to show that the currents in the F_1 region are substantially *in phase* with those producing the observed solar diurnal magnetic variations.

The magnitudes of the observed anomalies in $h'F_1$ appear to be of the order ± 4 km. Comparison with figure 1, which shows theoretical results for the E region ($H = 11$ km.), shows that this is of the order to be expected. In the F_1 region the increased value of H (30 km.), and the lower recombination rate, should give rise to $h'F_1$ deviations several times larger than those shown in figure 1, for the same drift velocity (22·5 km./hr.).

From what has been found in § 2, it might be anticipated that N_m also should show departures from the Chapman variation, which for this quantity is closely proportional to $\cos^{\frac{1}{2}}\chi$. Such deviations, of order 2 %, have been found. These deviations are obviously significant, being some 30 times greater than the probable observational error. Since their interpretation involves an additional parameter, the gradient of v, it has been thought best to leave their discussion to another occasion.

4. THE *E* REGION

Deviations from Chapman's theory in this region are likely to be smaller than those for the F_1 region for two reasons. The height of the homogeneous atmosphere in the E region is less than half that in F_1, and the rates of ion production and recombination are much larger for E than for F_1. For this reason it is desirable to use data from observatories which have been in operation at least 10 years. Unfortunately, the Huancayo and Watheroo observatories ceased to scale $h'E$ after 1 or 2 years' operation, on account of the smallness of the diurnal variation. This leaves Washington and Canberra as the only two observatories with 10 years or more available hourly data.

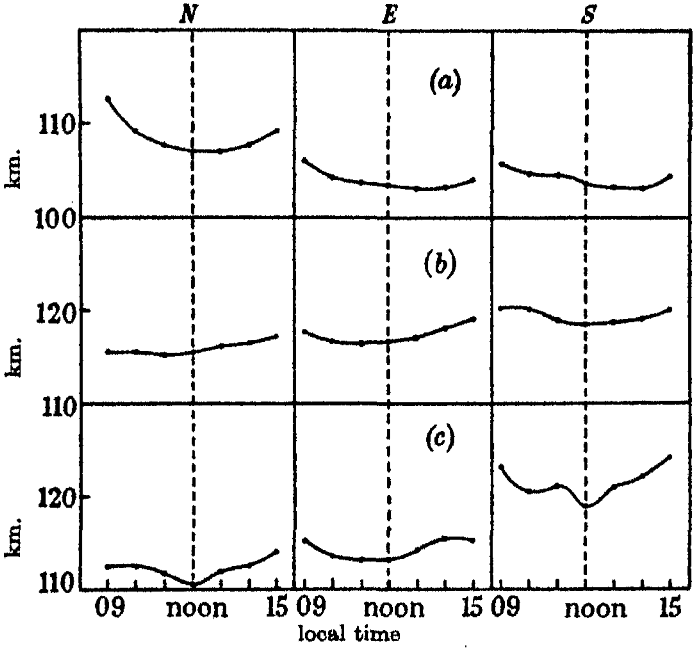


FIGURE 3. Average diurnal variations (09 to 15 hr.) of $h'E$ for each season (N, E, S) at (a) Canberra 1937 to 1948, (b) Washington 1934 to 1947, (c) Ottawa 1942 to 1947.

The mean diurnal variation of $h'E$ for each season at Canberra is shown in figure 3a for the years 1937 to 1948 inclusive. It will be observed that there is evidence of an upward deviation of the E region during the morning in all seasons (or downward deviation in the afternoon). This pattern is clear for the N and E months and slightly less so for the S months. Following the discussion of the previous section for the F_1 region, it might be anticipated that the current focus in the E region is south of

Canberra in all seasons, but is approaching the latitude of Canberra in the S months. If this be so then it would seem likely that Washington should show a reversed type of diurnal deviation. The corresponding Washington $h'E$ data for the period 1934 to 1947 are shown in figure 3*b*. It will be seen that the N and E months show morning depression and afternoon elevation, in accordance with the view that the current focus passes along a latitude between Washington and the equator. The diurnal variation of $h'E$ in S months at Washington is scarcely distinguishable from the Chapman variation.

Further confirmation of the position of the current focus is obtained from the study of six years (1942 to 1947) of $h'E$ hourly data from Ottawa (lat. 45·5° N, long. 75·8° W). It will be seen (figure 3*c*) that there is morning depression of $h'E$ at this site in all seasons. The plotted points in the figure lie less smoothly on a curve than the corresponding points for Canberra and Washington, thus showing the influence of the smaller amount of data available from Ottawa.

Following the reasoning outlined in §§ 1 and 2 it would appear to be a safe conclusion that the tidal electric forces have their maximum eastward direction in low latitudes at approximately 9 to 10 hr. local time. It follows that the currents produced by the solar tide in the E region are substantially in phase with the currents which produce the solar diurnal magnetic variations.

5. THE F_2 REGION

(5·1) *Foreword on methods to be used*

It is clear that the methods of analysis used above for the small deviations in the E and F_1 regions will not necessarily be best for the F_2 region. For the latter region the departures of N_m from Chapman's theory are gross, while the variations of h' and $h^{\text{max.}}$ bear no resemblance at all to those predicted by his theory. It has seemed best to attack first the N_m variation for two reasons: this bears at least some resemblance to Chapman's curves, and the analyses in II and III show that the lunar variations in f^0F_2 have less statistical scatter than those in $h'F_2$ and $h^{\text{max.}}F_2$. It seems reasonable to hope that if the variations in f^0F_2 can be satisfactorily accounted for by the action of the solar tide, then the meaning of the complexities in the h variations will become clearer.

The vast amount of ionospheric data accumulated in the last few years in almost all representative parts of the world shows that the F_2 region variations are complex, depending both on latitude and longitude (Appleton 1946). It would seem therefore to be difficult to select representative data without an elaborate harmonic analysis. This problem has been tackled by selecting for examination the data from four Australian ionospheric observatories, located at approximately the same longitude, but which cover a wide range of latitude. These are Cape York (lat. 11·0° S, long. 142·4° E), Brisbane (lat. 27·5° S, long. 153·0° E), Canberra (lat. 35·3° S, long. 149·0° E) and Hobart (lat. 42·8° S, long. 147·4° E). If the variations at these sites can be satisfactorily accounted for by the action of the solar tide, then it should not be difficult to extend the analysis to other longitudes, making due allowance for the additional term introduced by the earth's magnetic asymmetry about its axis of rotation (see I, § 4·4).

Before setting out the data and analysis, the principles to be followed in their interpretation are briefly enumerated. These have been discussed at length in the former papers.

(i) The F_2 region is distorted vertically by *semi-diurnal* electron drift produced by the solar tide.

(ii) The effect of this drift on f^0F_2 and hF_2 depends in sense and magnitude not only on the sense and magnitude of the drift, but also on the height gradient of the drift. (All these quantities are as yet unknown in the F_2 region for the solar tide.)

(iii) The effective electron recombination coefficient α in the F_2 region is much less than that in F_1, the half-life of F_1 ionization being of order minutes, while that of F_2 ionization is some hours (see I). It appears that α decreases rapidly above the $h^{\text{max.}}F_1$ level. It follows that the diurnal variation of f^0F_2 may be profoundly dependent on the phase of the semi-diurnal vertical drift. For example, suppose that near sunset the drift is zero near the level $h^{\text{max.}}F_2$, and is about to begin upwards. It follows that ionization then existing at that level will undergo, during the hours of darkness, a vertical oscillation which never carries it below that level. Its rate of disappearance during the night will be slow. On the other hand, if the drift is opposite in phase to that just postulated, then immediately after sunset ionization will move downwards into the region where α is larger and will disappear quickly. Again, in the daytime, the ionization may be expected to be great if at sunrise drift is about to begin upwards. In this case, if the level of N_m moves upwards with the drift, then the ionization there existing, and also that formed there during the day, will remain above the recombination 'sink'. Since the drift is semi-diurnal, it appears that the optimum conditions for large average values of N_m by day are the same as those for large night values. These general considerations may be summed up symbolically thus:

Large values of N_m (day and night): ↑9

Small values of N_m (day and night): ↑3

The arrow followed by a number is used here and below to indicate the hour at which the drift has its maximum upward (or downward ↓) velocity.

(iv) The drift may be expected to contain several harmonics, arising from the tidal velocity potential, as well as from the non-uniform and anisotropic conductivity of the F_2 region. The most important of these should arise from ψ_2^2 in the velocity potential. There may also be seasonal harmonics, the most important of which will probably be ψ_3^2. Global distribution of the drift arising from each of these harmonics has been shown in I (figure 5*a*, curves II and IV respectively). The term due to magnetic asymmetry which arises in the drift from ψ_2^2 is shown also in I (figure 5*b*). It acts in such a manner as to tend to make the global distribution of drift symmetrical about the magnetic equator, instead of about the geographic equator. Since the magnetic equator lies some degrees north of the equator in the Australian longitudes considered, the main effect of magnetic asymmetry will be to increase the effective latitude of each Australian observatory by a few degrees.

(5·2) *The diurnal variations of $N_m(F_2)$ in eastern Australia*

The average F_2 diurnal variations in N_m for each season are shown in figure 4 for the four sites in eastern Australia. Each of these observatories began to operate in different years, so that the data are not quite homogeneous. Thus Cape York (*a*) data is for 1946 to 1947, Brisbane (*b*) for 1943 to 1947, and Hobart (*e*) 1945 to 1947. To facilitate comparison two sets of curves are given for Canberra, one (*c*) for the period 1942 to 1944, which covers the years of sunspot minimum, and one (*d*) for the years 1945 to 1947. (It is remarked that these sets of curves would appear to be quite representative of the variations found in other parts of the world at corresponding magnetic latitudes. The only type of variation not exhibited by them is that found very close to the magnetic equator.)

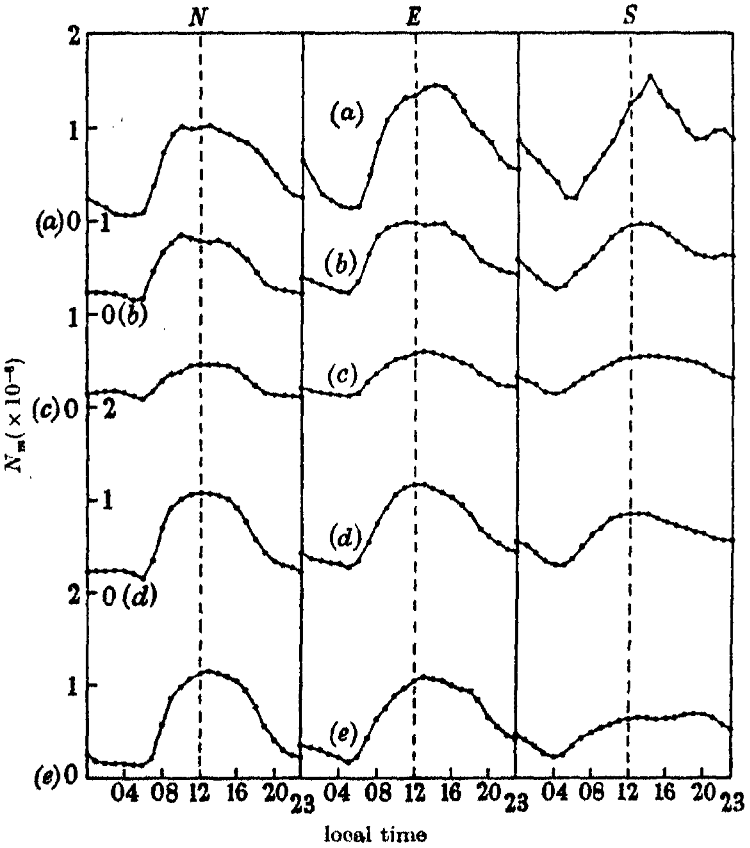


FIGURE 4. Average diurnal variations of $N_m(F_2)$ for each season (*N*, *E*, *S*) at (*a*) Cape York 1946 to 1947, (*b*) Brisbane 1943 to 1947, (*c*) Canberra 1942 to 1944, (*d*) Canberra 1945 to 1947, (*e*) Hobart 1945 to 1947.

In studying the curves in figure 4 attention is concentrated, as before, on the departures from the normal Chapman variation. For this region it is not so easy to decide on the norm, since α is low enough to give appreciable departure from equilibrium, and maximum N_m may occur an hour or more after noon. Nevertheless, it is clear that there are large departures from any possible norm. At low latitudes maximum N_m occurs before noon in N months, and after noon in S months; at high latitudes this maximum occurs in the afternoon in N and E months, and in the evening in S months. Other outstanding characteristics are the large winter values of N_m during daylight hours at high latitudes, and the low values at these latitudes in S months.

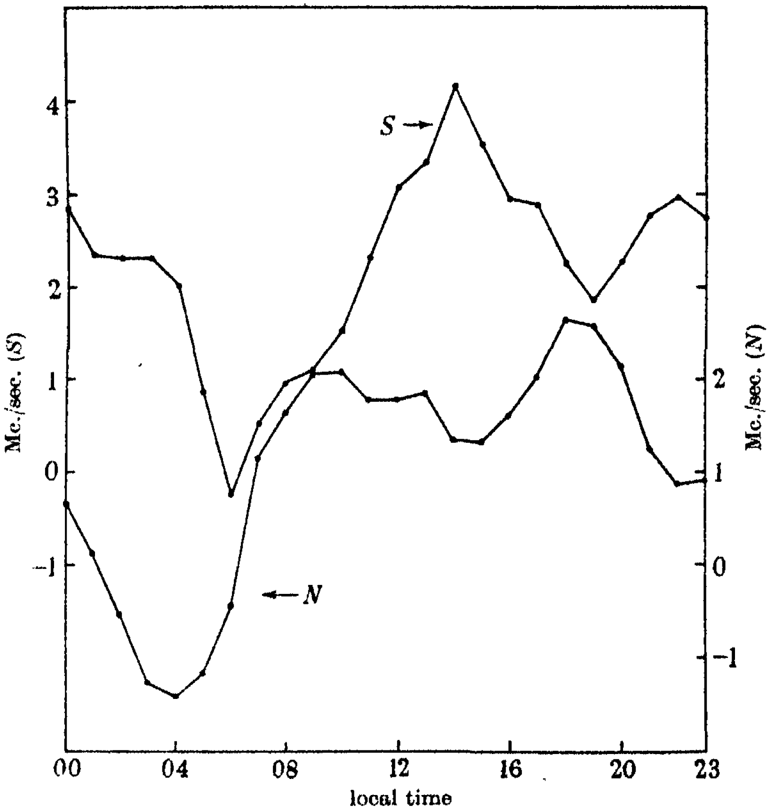


FIGURE 5. Differences between mean hourly values f^0F_2 at Cape York and Canberra for S and N months (1944 to 1946).

If the theory developed in I be valid then these characteristics should be explicable by a small number of 12-hourly surface harmonics. It is desirable then to examine the data for evidence of semi-diurnal harmonics. It is clear that these terms, if they exist, would become more obvious if the 'normal' diurnal variation could be even partially removed from the data. This can be achieved by subtracting the average diurnal variation at one site, for a given epoch, from that at another latitude for the same period. In this manner the Chapman variation will be partially eliminated, particularly when $\cos\chi$ is not greatly different at the two sites (i.e. in S months), while the difference in the semi-diurnal harmonics will remain. This procedure has been followed for the pair of stations Cape York and Canberra, for both N and S months, for the period 1944 to 1946 (figure 5). Care was taken to ensure that the averaged f^0F_2* data used were derived from identical days for each site. The same procedure was followed (figure 6) with data from Brisbane and Canberra for the period 1943 to 1947.

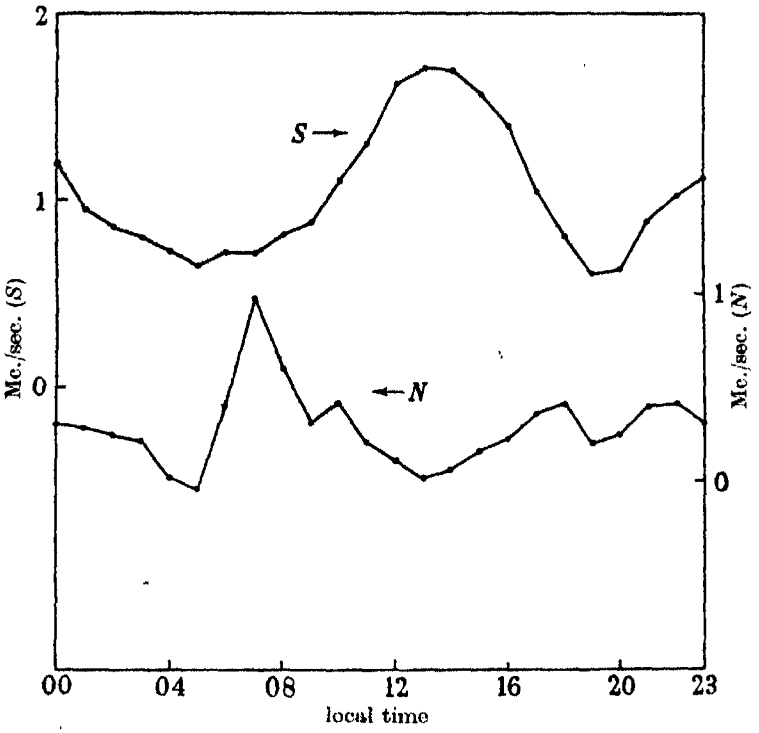


FIGURE 6. Differences between mean hourly values f^0F_2 at Brisbane and Canberra for *S* and *N* months (1943 to 1947).

It is clear that there is a large semi-diurnal harmonic in the differences of the diurnal variations at these sites. As is to be expected from the reasoning above, this harmonic is most clearly marked in *S* months. A surprising feature is the marked difference in phase of this harmonic in *N* and *S* months. This suggests that the seasonal semi-diurnal harmonic is comparable with ψ_2^2 in the F_2 region. Further evidence on this point has been obtained by the study of the equinoctial change in the semi-diurnal harmonic at one site. Figure 7 shows the difference between the average diurnal variation of f^0F_2 at Cape York for the period 7 to 20 March 1945 and that for the ensuing fortnight 22 March to 4 April. It is clear that a remarkably large change has occurred in the semi-diurnal harmonic at the equinox.

A series of studies using the above methods has suggested that it may be possible to account for the range of phenomena shown in figure 4 by the theory developed in I, making use principally of the two harmonics ψ_2^2 and ψ_3^2 in the tidal velocity potential. The global distribution of drift velocity due to these two harmonics has already been worked out (I, figure 5*a*, curves II and IV). Two distributions closely resembling these (figure 8) have been chosen for the purpose of interpreting the data. The small changes which have been made give a slightly better fit to the data, and are permissible on theoretical grounds, since it is known that higher harmonics must

---

* The critical penetration frequency f^0 is related to N_m by the equation $N_m = 1{\cdot}24 \times 10^4 (f^0)^2$, where f^0 is measured in Mc./sec.

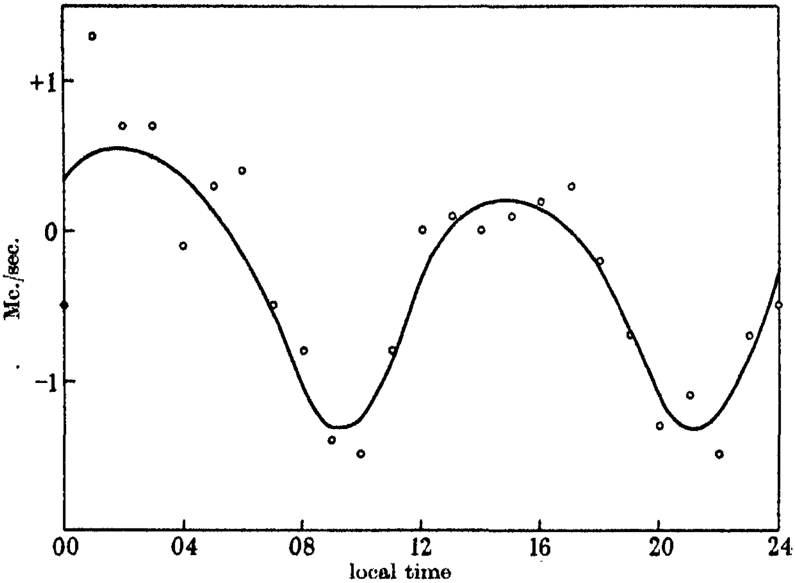


FIGURE 7. Difference between average diurnal variation f^0F_2 at Cape York for 7 to 20 March 1945 and same quantity for 22 March to 4 April 1945.

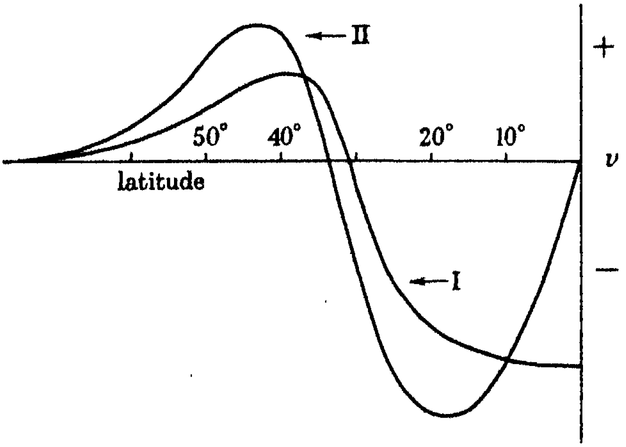


FIGURE 8. Assumed latitude distribution of vertical drift velocity v (arbitrary units) arising from semi-diurnal sectorial harmonics (curve I), and from semi-diurnal tesseral (seasonal) harmonics (curve II).

be present. Curve I in figure 8 gives the assumed distribution of drift velocity arising from the semi-diurnal sectorial harmonics, and curve II that from the semi-diurnal tesseral (seasonal) harmonics. It is now necessary to assign phases to these harmonics. For this purpose it is assumed, in accordance with previous findings, that the variation in N_m is proportional to the drift velocity, and bears a definite but as yet unknown phase relationship to it. Figure 8 can now be regarded as giving the global distribution of the N_m variations, and phases can be assigned to curves I and II in the attempt to account for the global distribution of the observed variations. These phases have been assigned as follows. The diurnal variation of N_m at Cape York at the equinoxes shows a clear midnight maximum (figure 4), being some 15 % greater at 24 hr. than at 23 hr. Accordingly, the N_m variation given by curve I is assigned maximum positive amplitude at 12 hr., for latitudes lower than 31° (and consequently at 6 hr. for higher latitudes). Curve II is assigned its maximum phase at 3 hr. in summer for latitudes below 33·5° (and consequently at 9 hr. for higher latitudes). In winter the phase of curve II reverses, of course, giving maximum N_m at 9 hr. in low latitudes, and 3 hr. for high latitudes. Having assigned these phases on the basis of the observed variations at Cape York, it is now possible to test their applicability to other latitudes. To do this, the vectors represented by curves I and II (figure 8) are added, the result being displayed on the harmonic dial (figure 9). The curves (N and S) give the amplitude and phase of the N_m variation at all latitudes for N and S months. The scale of latitude is shown along each curve. At a given point on the curves the amplitude of the N_m variation is given by the distance from the centre

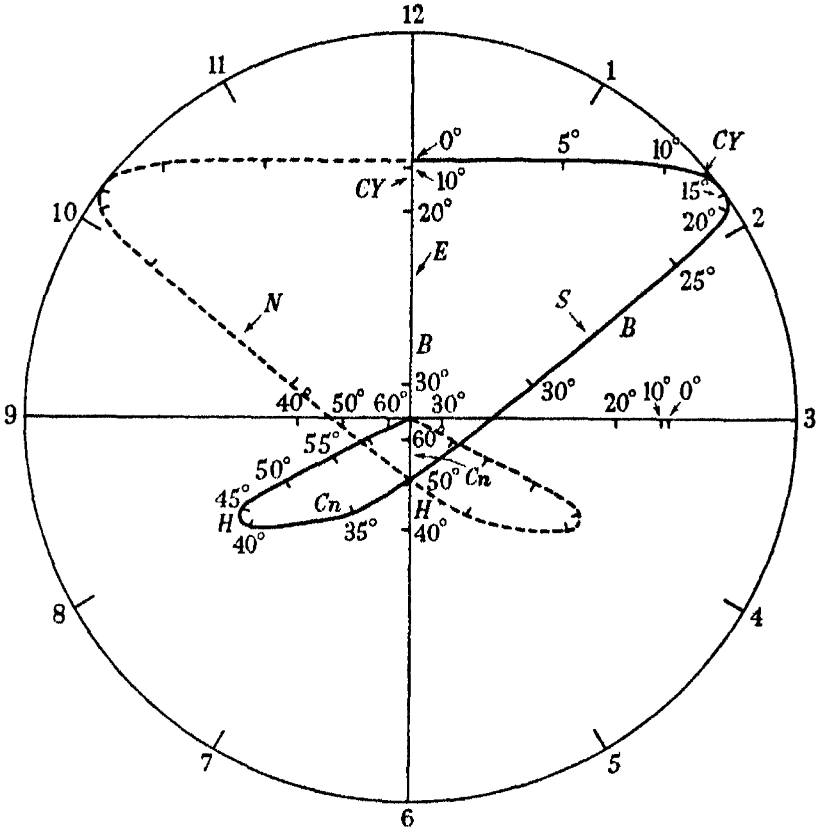


FIGURE 9. Harmonic dial giving amplitude and phase of semi-diurnal tidal N_m variation for each season (N, E, S) at all latitudes.

of the dial, while the phase (time of maximum positive amplitude) is given by drawing a straight line from the centre of the dial through the point; this line is then considered the hour hand on a 12 hr. clock face. The corresponding curve for the equinoxes is simply the vertical line through the centre of the dial. It too is scaled in latitude. The position of the Australian observatories is shown on each curve by the appropriate initials, CY, B, Cn, H. Comparison of the results in figure 9 with the observed variations in figure 4 would appear to show that a good representation of the N_m variations has been obtained. Thus, at Hobart, the S maximum at 20 hr., the E maximum at 18 hr., and the N elevations at 4 and 16 hr. are all accounted for. Similar features at the appropriate times can be traced in the observed data for the other observatories.

There remains to be considered the relationship between the phase of the N_m variation and the drift which produces it. An important clue is provided by the low summer-day values of N_m found at latitudes $> 40^\circ$ in all parts of the world. According to the discussion in § 5·1 this should happen* when the phase of the drift is approximately $\downarrow 9$. Similarly, a high degree of winter-day ionization at latitudes $> 40^\circ$ strongly suggests drift which has phase near $\downarrow 3$. Now these phases are approximately observed in the N_m variation at high latitudes, which has its maximum at about 8 (and 20) hr. in summer, and 4 (and 16) hr. in winter. It would appear very probable, therefore, that the maximum downward drift occurs very nearly at the same time as maximum positive δN_m. (It will be recalled that this is the phase relationship which has been found between these quantities for the case of the lunar tide (II and III). Theoretically, such a phase relationship has been found to indicate that the phase of the drift velocity *advances* upwards through the ionized region.)

Now, if this interpretation be correct, then the drift velocity at low latitudes in winter will have the phase $\downarrow 10$, and low values of daytime ionization might be anticipated there in that season. There is no evidence that this happens. A good deal of light is thrown on the cause of this apparent difficulty when the phase of the drift in the F_1 region is considered in conjunction with that in F_2. It appears (§ 2) that the phase of the F_1 drift is much less influenced by seasonal harmonics, and can be represented roughly by $\uparrow 3$ in high latitudes ($> 35^\circ$) and $\uparrow 9$ in low. Curve I in figure 8 can probably be taken as an adequate representation. In figure 9 it appears as the horizontal line through 9 and 3 hr., and has been scaled in latitude. The rapid change of amplitude and phase between 30° and 40° latitudes is clearly seen; Canberra is therefore particularly sensitive to even a small seasonal harmonic, as the evidence in § 2 shows.

It will be seen that at both high and low latitudes in summer, the phases of the drifts in the F_1 and F_2 regions agree to within an hour or so; in winter there is near-phase opposition in all latitudes. (It is also to be remarked that in nearly all latitudes, at both seasons, the phase of the F_2 drift leads that of the F_1, thus exhibiting the upwards advance of phase which is found above on other evidence.) Now a phase reversal, such as appears in winter between F_1 and F_2, implies a small amplitude at the level of reversal. It follows that, in winter at low latitudes, the F_2 drift in the morning hours before noon, although downwards, has a small and decreasing amplitude. This cannot be so effective in removing the under-surface (and maximum) of the F_2 region as the corresponding drift in high latitudes in summer, which sweeps downwards through both regions unchecked. The 'blow-up' of the F_2 region in summer at high latitudes, which is one of the most outstanding characteristics of this region, has been ascribed above to downward drift of ionization in the hours between sunrise and noon. This process is irreversible, since it results in the disappearance of much of the ionization formed in the lower part of the regions. While it is easy to understand the low values of daytime N_m which accompany this process, it is not at all obvious why it is accompanied by high values of $h'F_2$ and $h^{\text{max.}}F_2$, which are the other characteristic features of the 'blow-up' of the region. It is not easy to treat the problem analytically, in view of the complexity of the relevant differential equation, but some insight may be obtained from cases already examined in I. The increasing value of α at the lower levels results in an increasing rate of disappearance of electrons as we pass downwards in the region. This effect is simulated, for the simple cases of variable flow treated in I, by the condition that the velocity increases in amplitude as we pass downwards. In the latter case the gain in ionization per unit volume per second from above is less than the loss to lower levels; there is a net loss from flow alone. It follows that the case of uniform downward flow accompanied by a gradient of α positive downwards, should be closely analogous to that of uniform α with drift increasing downwards. The latter case has been solved in I (figure 6). It is shown there that when the gradient of v is sufficiently steep $h^{\text{max.}}$ rises. There is little doubt that the gradient of α is sufficiently steep to give the corresponding effect in the case under consideration here, and that drift downwards will produce the characteristic 'blow-up' of the F_2 region. There is one fundamental difference, however, between the two cases just compared. That treated in I is reversible; when the drift reverses, and ionization moves upwards, $h^{\text{max.}}$ will decrease and N_m increase. The present case is irreversible; much of the ionization carried downwards has disappeared into the 'sink', and on reversal of the drift $h^{\text{max.}}$ will continue to increase, though N_m may now start to increase. It will be seen that, for the conditions postulated, the level at which N_m occurs will move upwards, whether the drift be up or down. Now at sunrise these conditions appear to be always satisfied. The F_1 and F_2 regions at this time are merged together, and the sharp gradient in α must be close to the level of $h^{\text{max.}}F_2$. It follows that, under all conditions of appreciable drift at sunrise, the F_2 region will commence to rise and separate from F_1. But from what has been said above, such drift is appreciable always in summer at all latitudes. In winter at low latitudes the small downward drift soon after sunrise may cause a small 'blow-up', but in winter at high latitudes the small upward drift may fail to separate the regions appreciably. The ionization formed by the sun is slowly lifted from the sink, and thus preserved to contribute to the high midday level of ionization which is so characteristic of high latitudes in winter.

---

* This depression of N_m is irreversible, and is to be clearly distinguished from the cyclic variation of N_m portrayed by figures 8 and 9.

The above discussion is clearly consistent with the view, already expressed by Bradbury (1938) and Mohler (1940), that the F_1 and F_2 regions can be regarded fundamentally as manifestations of a single Chapman region. The greater electron density in F_2 can be regarded as due to the lower value of α in that region. On this view the reduction of α must occur slightly above the level of maximum ion production in this single region, which is to be identified with the level of $h^{\max.}F_1$. This is consistent with a conclusion of Appleton (1940), who finds α constant over the small range of heights covered by the diurnal and seasonal changes in $h^{\max.}F_1$. On this view the very large changes in f^0F_2 with sunspot activity are due to a component of solar radiation which is quite distinct from that producing the normal composite F region. It would be anticipated then that this highly variable component would be less penetrating than the normal component, and would produce its maximum ionization at somewhat higher levels in F_2. There is a good deal of evidence that this is so, but the discussion of the data on this point must be left to a later paper.

None of the Australian observatories lies close to the magnetic equator, so that the characteristics of the region have not been discussed so far. The effect of the term in ψ_2^2 due to magnetic asymmetry of the earth is to distort slightly curve I in figure 8, so that it reaches its lowest point, not at the equator, but near to the magnetic equator. For the purposes of this discussion it is sufficient to assign to such observatories a latitude which is their magnetic and not their geographic latitude. It follows from figure 9 that the phase of the drift at such observatories will be ↓ 12, and the seasonal variation will be small. The amplitude of the drift will be considerably larger than is shown in figure 8.

6. THE LOCATION OF THE CURRENTS PRODUCING THE MAGNETIC VARIATIONS

It has been shown above that the 'dynamo' currents flowing in the E and F_1 regions are substantially in phase with those which produce the solar magnetic variations. The corresponding currents in the F_2 region show marked seasonal variations of phase which do not appear in the magnetic variations. There is little doubt therefore that the currents in the F_2 region have only a small effect in producing the magnetic variations, a conclusion which is *a priori* probable because of the small conductivity of this region. Now McNish (1937) has produced strong evidence that the solar magnetic variations are produced mainly in the D region. This conclusion is based on the fact that notable solar flares, which produce strong ionization in this region alone, produce also a strong enhancement of the normal solar magnetic variation. It appears therefore that this is the main seat of the solar magnetic variation, while the currents in the E and F_1 regions contribute to an unknown extent.

On the other hand, it was shown in III that the lunar 'dynamo' currents in the latter regions (and in F_2) are almost in phase opposition to those which produce the lunar magnetic variations. Symbolically, the evidence from the study of solar and lunar drifts in the ionosphere indicates that

$$S = S_D + S_{EF}, \tag{2}$$

while

$$L = L_? - L_{EF}, \tag{3}$$

$L_?$ being the main lunar magnetic variation current. On theoretical and observational grounds there is little likelihood of any unknown ionized region existing which could have a conductivity adequate to produce observable magnetic variations. It seems, therefore, almost essential to identify $L_?$ as L_D.

Now it has been a serious difficulty in the theory of the magnetic variations that no single ionospheric region has the necessary properties to permit identification as the seat of the L variation, which shows large seasonal changes, but almost negligible influence from sunspot activity. It is important, therefore, to investigate whether the system given by equation (2) is capable of accounting for these peculiarities of the lunar variation.

The ionospheric current systems necessary to produce the solar and lunar magnetic variations have been delineated by Chapman & Bartels (1940, 1, 229 and 262 respectively). For S the daytime circuit in winter carries 36,000 A, in summer 89,000 A. For L the daytime circuit carries 10,000 A in summer and 2400 A in winter. These figures refer to years of sunspot minimum. At sunspot maximum the S currents increase by 50 %; the L currents are unchanged. These facts are expressed symbolically in terms of equations (1) and (2) as follows:

$$1+\rho = 36/S_D, \tag{4}$$

$$1-\rho = 2{\cdot}4/kS_D, \tag{5}$$

$$p_1+\rho p_2 = 89/S_D, \tag{6}$$

$$p_1-\rho p_2 = 10/kS_D, \tag{7}$$

$$q_1+\rho q_2 = 54/S_D, \tag{8}$$

$$q_1-\rho q_2 = 2{\cdot}4/kS_D, \tag{9}$$

where S_D is the solar daytime current in the D region circuit in winter, measured in thousands of amperes, ρ is the ratio of the conductivity of the EF regions to that of the D, k is the ratio of the average lunar tidal velocity to that of the solar, $p_{1,2}$ are the ratios of summer to winter conductivities in the D, EF regions, $q_{1,2}$ are the ratios of sunspot maximum to sunspot minimum conductivities in the D, EF regions.

It is assumed for simplicity that both the solar and lunar velocities are constant in magnitude throughout the regions, although, of course, the latter velocity changes its sense between D and EF.

In this set of six equations there are seven constants, one of which, q_2, can be derived from the variation of electron densities in the E and F_1 regions during the sunspot cycle. On the assumption that the proportion of negative to positive ions in these regions does not change during this cycle q_2 may be taken as 1·6. Equations (4) to (9) are now solvable, giving $S_D = 2{\cdot}1\times10^4$ A, $\rho = 0{\cdot}71$, $k = 0{\cdot}40$, $p_1 = 2{\cdot}7$, $p_2 = 2{\cdot}1$, $q_1 = 1{\cdot}43$. All these results seem entirely reasonable; particularly so when it is realized that there is nothing to prevent some negative solutions being obtained from such a system of equations, a result which would not have been acceptable physically.

It appears then that the facts of the S and L variations can be interpreted naturally by considering each to have two unequal components, which are additive for the S variation and which *oppose each other in the L variation*, this interpretation being consistent with the evidence derived above on the phases of the solar and lunar tides in the EF region. Since this conclusion has been reached indirectly, and depends upon a rather long chain of evidence and deduction, it is clearly desirable to submit it to a direct test. It appears that this should be possible from studies of the effects of solar flares upon the magnetic variations. According to the solutions found above L_D/S_D is as much as 0·4, although the observed L variation is only 2·4/36 = 0·067 time S. Clearly, when solar flares enhance the conductivity of the D region, they should greatly enhance L, so that it becomes more comparable with S. In fact, it should be possible to find evidence of a greatly enhanced L in data such as McNish (1937) studied. The latter has given analyses for three major flares, in the form of world maps. Arrows from each magnetic observatory indicate (*a*) the normal ionospheric current vector above the observatory at the time of the flare, and (*b*) the additional current vector due to the flare. McNish concluded that these vectors coincided substantially in direction. A careful re-examination of this data appears to support the view that the deviations between these two vectors, as shown on McNish's maps, agree with the hypothesis that they are due to an enhancement of L. McNish's data for Huancayo are reproduced in figure 10.

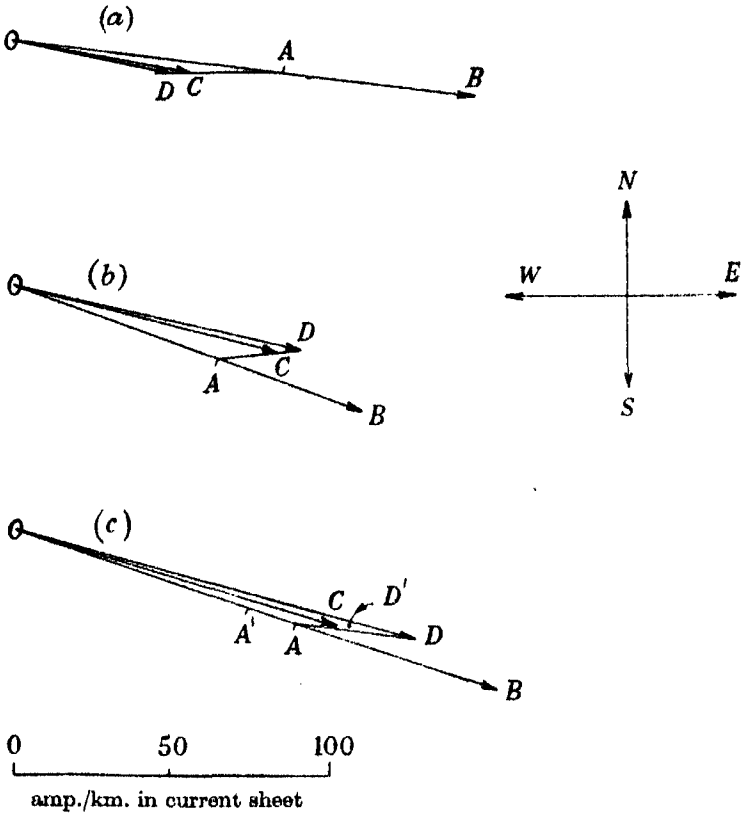


FIGURE 10. Ionospheric current sheet vectors at Huancayo during three solar flares. **OB** is normal diurnal vector, and **OC** the additional vector due to flare (according to McNish). **AD** is predicted enhanced lunar vector.

This site was chosen for special study since it was near the subsolar point on the occasion of each flare, and both S and L are particularly large at this observatory. In the figure the current vectors are shown at (*a*) for the flare of 6 November 1936 at 1609 G.M.T., at (*b*) for that of 25 August 1936 at 1825 G.M.T., and at (*c*) for 8 April

1936 at 1646 G.M.T. In each case OB gives the direction and amplitude of the current producing the normal diurnal variation at Huancayo at the given time, according to McNish. OC gives the additional current at the time of the flare, also according to McNish. OA gives that portion of the S currents which flow in the D region, according to the above analysis, i.e. $1{\cdot}71\ \mathbf{OA} = \mathbf{OB}$. If S only were enhanced by the flare, then **OC** should coincide in direction with **OB**. The direction of L at the date and time of each flare has been derived from Chapman & Bartels (1940, **1**, 262–263). It is given in each case by the direction AD. It is necessary now to assess the enhancement of conductivity in the D region due to each flare. If this can be done, the additional vector due to both S and L is fully determined by the theory above. For a first attempt it has been assumed that the conductivity of the D region was doubled by each flare. **OA** then represents in each case the enhancement of the S current, while the theoretical enhancement of the L current has the direction AD and magnitude $0{\cdot}4\ OA$. Thus in each case **OD** represents the predicted total flare disturbance vector. This is to be compared with **OC**, that actually found by McNish, and **OA**, that which S alone would produce. It is seen that there are grounds for the view that a greatly enhanced L variation appears in the flare disturbance vector. In each case the S vector has been deviated by L in the right direction. In cases (*a*) and (*b*) the magnitude of the change of conductivity in the D region appears to have been well assessed, so that the calculated vector agrees closely with that observed. In case (*c*) the magnitude of the total disturbance vector would agree more closely with that found if a slightly smaller enhancement of D region conductivity occurred for this flare than for the other two. Thus if **OA′** represented the enhancement in S in this case, then **OD′** would be the deduced vector. The study of the effects of other major flares on the Huancayo magnetic variations, particularly if they occur at times when the L currents are at right angles to the S, should decide the issue conclusively.

This work has been done on behalf of the Radio Research Board of the Council for Scientific and Industrial Research, Australia. It has been carried out principally at the Commonwealth Observatory. The thanks of the author are due to the Commonwealth Astronomer for the provision of facilities and ionospheric data, and for useful discussion. The statistical work was carried out mainly by Miss M. Kilmartin and a group of computers supervised by Miss B. Hardwick. The author is indebted to various institutions, notably the Department of Terrestrial Magnetism, Carnegie Institution of Washington, for much of the data used.

REFERENCES

Appleton, E. V. 1940 *Trans. Wash. Meet. Ass. Terr. Magn. Elect.* p. 523.
Appleton, E. V. 1946 *Nature*, **157**, 691.
Bradbury, N. E. 1938 *Terr. Magn. Atmos. Elect.* **43**, 55.
Chapman, S. 1931 *Proc. Phys. Soc. Lond.* **43**, 26.
Cowling, T. G. 1932 *Mon. Not. Roy. Astr. Soc.* **93**, 90.
McNish, A. G. 1937 *Terr. Magn. Atmos. Elect.* **42**, 109.
Martyn, D. F. 1947 *Proc. Roy. Soc.* A, **189**, 241.
Martyn, D. F. 1947 *Proc. Roy. Soc.* A, **190**, 273.
Martyn, D. F. 1948 *Proc. Roy. Soc.* A (in the Press).
Mohler, F. L. 1940 *Bur. Stand. J. Res., Wash.*, **25**, 507.

Temperature and pressure phenomena in the flow of saturated liquids

By R. S. SILVER, D.Sc.

(*Communicated by Sir Alfred C. Egerton, Sec.R.S.—Received* 12 *November* 1947—
Revised 11 *March* 1948)

The rate at which evaporation can occur from a surface is limited by the rate of heat flow into the surface. Because of this limitation when a liquid at its boiling point commences to flow, the proportion of vapour which forms as a result of the Bernoulli fall in pressure is less than is calculated from adiabatic expansion. A theory for the actual amount of vapour which will form is given, together with an analysis of its effect upon the velocity of flow and the quantity of fluid passed by a nozzle. It is found that although the system consists of a central core of liquid and an envelope of vapour, the calculated limiting velocity of flow, which is the velocity of pressure propagation in the system, is low. The order of magnitude is 20 m./sec. as compared with 400 and 1500 m./sec. for vapour alone and liquid alone respectively. There is a corresponding critical pressure, analogous to that in a gas nozzle, but in contrast to the gas case, both the critical pressure and the limiting velocity itself are functions of the diameter and length of nozzle.

Experiments are described which confirm these theoretical deductions in all respects, and which also show good quantitative agreement with calculated velocities, critical pressures, and amount of discharge and of evaporation.

On the basis of this experimental verification general phenomena to be expected in pressure propagation through liquids near saturation temperature are discussed.

1. INTRODUCTION

Consider a volume of liquid at a uniform temperature T_0 being introduced into an atmosphere of its own vapour in the saturated condition at pressure p and temperature T. If $T_0 > T$ the whole volume of liquid is instantaneously in a superheated condition, its temperature being greater than the saturation value corresponding to the pressure upon it. It will therefore proceed to evaporate into the surrounding atmosphere until such a proportion has evaporated that the remaining liquid is reduced to temperature T and the whole system of liquid and vapour is in saturation equilibrium at pressure p. It is assumed, of course, in enunciating these conditions that the conditions p, T in the vapour atmosphere are maintained by suitable experimental arrangement. The proportion of the original liquid volume which has to evaporate for the new equilibrium is readily calculated on the following considerations. Let L be the heat of vaporization, C_l the specific heat of the liquid, M the original mass of the liquid and q the proportion which evaporates, then we have

$$qML = C_l(T_0 - T)M. \tag{1}$$

This expression is an approximation valid when $T_0 - T$ is small. The more correct statement is obtained by considering that the entropy of a mixture of liquid and vapour with proportion q of vapour is $S = S_l + q(L/T)$. When this is differentiated and it is assumed that no heat is lost or gained in the complete system liquid plus vapour

$$dq = -\frac{dT}{L}\left[C_l + qT\frac{d}{dT}\left(\frac{L}{T}\right)\right].$$

From the Clausius-Clapeyron relation, where C_s is the specific heat of the saturated vapour, this gives

$$dq = -\frac{dT}{L}[C_l + q(C_s - C_l)]. \tag{2}$$

It is seen from this that equation (1) is an approximation valid when q is small. For larger values equation (2) would have to be integrated.

However, the results expressed either in equation (1) or in the integral of equation (2) refer to the attainment of a new equilibrium. The interest of the present paper is in circumstances where the time is too short for that equilibrium to be attained. The order of magnitude of the time required for obtaining equilibrium may be investigated as follows. The mass of liquid can be regarded as a body initially at uniform temperature T_0 which is suddenly introduced into a region into which it loses heat from its surface. It is an essential characteristic of the solutions to this type of problem that the mean temperature in the body changes appreciably only after a time which is of the same order of magnitude as $\lambda^2/5k$. In this expression k is the thermometric conductivity of the substance, i.e. its thermal conductivity divided by its specific heat per unit volume, and λ is a length of the same order of magnitude as a radius vector from a median point within the body to the boundary surface. Thus for a sphere of radius a losing heat at its surface, λ is of order a and is so also for a cylinder of radius a losing heat from its curved surface. Hence the new equilibrium postulated can only occur after a time of the order $\lambda^2/5k$. Now for most liquids k is of the order 10^{-3} cm.^{2}/sec., so that if λ is of the order 1 cm. the attainment of equilibrium will require more than 200 sec.

There is, however, a modification to this simple statement which has to be considered. If discontinuities can arise within the liquid its outer surface can no longer be regarded as definitive and equilibrium may be reached in a shorter time. Conversely, the possibility of equilibrium being attained in times shorter than $\lambda^2/5k$ is governed by the possibility of forming discontinuities within the liquid, and the actual time required will depend upon the number of such discontinuities formed. If it were true that a liquid whose vapour pressure is higher than its hydrostatic pressure could forthwith form an even distribution of bubbles throughout its volume, then the surface available for evaporation would be very much increased and the time required for equilibrium could become very short and approach zero. However, it has long been established theoretically that there is considerable difficulty in forming a bubble due to the action of surface tension, and, indeed, from an elementary consideration of surface tension it is more difficult to understand why bubbles do form in common practice than why they do not under certain circumstances. An amount of experimental evidence is now available which establishes that when bubbles do form it is because of the presence of solid irregular surfaces in contact with the liquid or of gas dissolved in the liquid or of mechanical tensile stress applied to the liquid, so that the effective hydrostatic pressure is much less than the vapour pressure. Thus it is legitimate to assume for a pure liquid not in contact with solid surfaces and not subject to tensile stress that the theoretical impossibility of forming bubbles will be fulfilled and therefore that the time required for attainment of equilibrium will be of the order discussed above.

This circumstance has many applications of practical importance. Thus, for example, when it is desired to cool a liquid by self-evaporation from a falling spray it is necessary that the time for fall of a drop should be of the order $a^2/5k$, where a is the radius of the drop. Such considerations determine the useful particle size. Another interesting case is where liquid at its saturation temperature flows through an orifice or short nozzle. It may be considered to flow as a jet, and in most practical circumstances it is actually in the orifice or nozzle for only a fraction of a second. Hence if the jet radius is of the order of a centimetre the attainment of equilibrium in an element of the liquid is not possible until after it has passed through the nozzle. It is with this problem in particular that the present paper is concerned, for besides being one of important practical occurrence it has also many features of considerable physical interest. One of the most important is that the acquirement of velocity in the orifice is accompanied by the formation of vapour. The specific volume of vapour is so large compared with the specific volume of liquid that even a very small proportion of evaporation may cause an appreciable change in the specific volume of the fluid as a whole. This means that the mass which can flow through the orifice, whose area is limited, will be less than the mass of cold liquid which could flow through the same orifice with the same velocity. However, it has also to be remembered that the velocity cannot be the same as for a cold liquid, since the specific volume and therefore its integral over the fall in pressure has been changed. Hence the calculation of the amount of saturated fluid which can flow through an orifice or nozzle requires the determination of three main things:

(1) the proportion of vapour which can form in the time during which the fluid is in the orifice;

(2) the effect of this proportion of vapour upon the specific volume; and

(3) the effect of this proportion of vapour upon the velocity.

However, all these three factors are interconnected, since the proportion of vapour depends upon the time in the nozzle and the time depends upon the velocity. The latter in turn depends upon the proportion of vapour formed and upon its effects on the specific volume. Moreover, the vapour will form an annular envelope around the liquid core, and the surface from which evaporation can take place will be determined by the radius of the liquid core rather than the radius of the nozzle.

However, the importance of the time necessary for attainment of equilibrium in this problem has not always been appreciated, and a conventional approach to the problem has been to assume that the proportion of vapour formed was given by the equilibrium formula of equations (1) or (2). This meant, of course, a greatly simplified procedure, since the proportion of vapour was then independent of the velocity and of the jet dimensions and could indeed be immediately specified for any fall in pressure. The velocity and the specific volume could then be calculated directly and an expression obtained for the quantity discharged. But naturally enough the results predicted were not at all in accordance with experiments. Thus Bottomley (1937) found that orifices draining saturated water from a steam boiler feed heater passed more than five times the quantity he calculated in the equilibrium theory. Benjamin & Miller (1941) found that sharp-edged orifices discharged the same amount of saturated water as of cold water. The latter result is immediately acceptable if one regards the time available for vapour formation to be zero for a sharp-edged orifice. No experiments have been published in which the problem has been dealt with from the standpoint of the time required for attainment of equilibrium except an account (Silver & Mitchell 1945) of the present experiments given from the engineering point of view, and in relation to which this theoretical discussion is presented.

2. THE VELOCITY OF FLOW OF A SATURATED MIXTURE

Consider a liquid originally in equilibrium at p_0 and T_0. Now let it flow without friction to a region of reduced pressure p_a whose associated saturation temperature is T_a. We have already noted that equilibrium conditions may not be attained in the short time available for flow so that the proportion of vapour which forms is not predictable from equilibrium considerations alone. Hence, denoting this proportion by q we cannot assume that q is proportional to $T_0 - T_a$. But while this is so, it is nevertheless legitimate to assume as an approximation that the fraction x/q which has formed at any intermediate temperature T is proportional to the temperature drop $T_0 - T$. Then since $x = q$ when $T = T_a$ the coefficient of proportionality must be $q/(T_0 - T_a)$. Therefore

$$x = \left(\frac{T_0 - T}{T_0 - T_a}\right) q. \tag{3}$$

This is the proportion of vapour at any intermediate pressure p and its associated T. The specific volume is

$$\begin{aligned} V &= V_l + x(V_v - V_l) \\ &\doteqdot V_l + xV_v, \end{aligned} \tag{4}$$

where V_v is the specific volume of the saturated vapour at p and T and V_l is that of the liquid.

The velocity is given by Bernoulli's theorem

$$\begin{aligned} d\left(\frac{v^2}{2}\right) &= -V\,dp \\ &= -V_l dp - xV_v dp. \end{aligned}$$

The integration of the first term is easy, since V_l is constant. Since V_v and T are saturation values, for dp in the second term can be substituted from Clapeyron's equation

$$\frac{dp}{dT} = \frac{L}{T(V_v - V_l)} \doteqdot \frac{L}{TV_v}, \tag{5}$$

and, using also (3), $$xV_v dp = \frac{T_0}{T_0 - T_a}\left(\frac{qL}{T}\right) dT - \left(\frac{qL}{T_0 - T_a}\right) dT,$$

therefore $$d\left(\frac{v^2}{2}\right) = -V_l dp - qL\left[\frac{T_0}{T_0 - T_a}\frac{dT}{T} - \frac{dT}{T_0 - T_a}\right]. \tag{6}$$

Variation of L with T can be neglected in the range of our experiments. The limits are from p_0, T_0 to p_a, T_a, and the initial velocity is taken to be zero. Hence the velocity obtained is

$$\frac{v^2}{2} = V_l(p_0 - p_a) + qL\left[\left(\frac{T_0}{T_0 - T_a}\log\frac{T_0}{T_a}\right) - 1\right].$$

If the temperature difference $T_0 - T_a$ is small compared with T_0 or T_a, these being absolute,

$$\left(\frac{T_0}{T_0-T_a}\log\frac{T_0}{T_a}\right)-1 \doteqdot \frac{T_0-T_a}{2T_a},$$

therefore

$$\frac{v^2}{2} = V_l(p_0-p_a)+\frac{qL(T_0-T_a)}{2T_a}. \tag{7}$$

It is worth noting that the first term is the expression for the velocity energy obtainable by liquid alone.

There is perhaps some difficulty in accepting this treatment in general since the fluid is not homogeneous. It may be supposed to divide itself into regions such as a liquid core and a vapour envelope, and through these regions stream-lines might be drawn in which the velocity was governed only by the pressure-volume relation of the separate phases, the velocity in the vapour phase being different from the liquid. In the present treatment this possibility is ignored. Theoretically it is only an approximation. As will be seen later the procedure is justified by results.

Equation (7) can be written in the form

$$v_a^2 = v_c^2\left(1+\frac{qL(T_0-T_a)}{T_a v_c^2}\right), \tag{8}$$

where v_c is the velocity acquired by cold water for the same fall in pressure, and this is of the form

$$v_a^2 = v_c^2(1+mq), \tag{9}$$

where

$$m = \frac{L(T_0-T_a)}{T_a v_c^2}. \tag{10}$$

m is a constant for given temperatures and pressures, so that (9) is an equation defining v_a in terms of q, i.e. which specifies the velocity in terms of whatever proportion of vapour does actually form.

3. TEMPERATURE DISTRIBUTION IN THE LIQUID CORE

Imagine the cylindrical jet at initial temperature T_0 divided into elementary circular laminae of radius a with no heat flow across the flat faces. For the heat flow in such a lamina, if θ is the temperature excess above the surrounding environment,

$$\frac{\partial^2\theta}{\partial r^2}+\frac{1}{r}\frac{\partial\theta}{\partial r}=\frac{1}{k}\frac{\partial\theta}{\partial t}. \tag{11}$$

The boundary conditions on entering an atmosphere at temperature T' are expressed in

$$\theta = T_0 - T', \quad t=0, \quad 0 \leqslant r \leqslant a, \tag{12}$$

$$\theta = 0, \quad t = 0, \quad r = a. \tag{13}$$

The solution is

$$\theta = 2(T_0-T')\sum_1^\infty \frac{J_0(R_n r/a)}{R_n J_1(R_n)}\exp\left[-\frac{R_n^2 kt}{a^2}\right], \tag{14}$$

where J_0 and J_1 are the zero- and first-order Bessel functions, and R_n is the nth root of the equation $J_0(x) = 0$.

The initial temperature distribution is as shown in figure 1*a*. Equation (14) gives for small values of kt/a^2 a temperature distribution as in figure 1*b*.

The values of kt/a^2 which are involved in flow through nozzles are very small, of order 10^{-4} to 10^{-3}. It is difficult to calculate the total heat lost from a lamina for such a short time, because one has to use very many terms of the series for θ. Moreover, in the present case the time considered is that for passage through the nozzle, and depends upon the velocity—which in turn is determined by the heat lost in evaporation. As a result of these two circumstances the problem is intractable on the strict basis of (14).

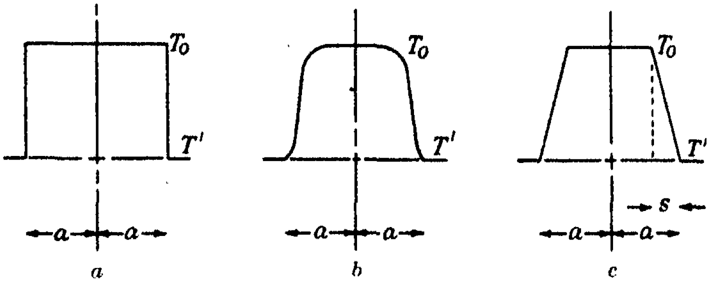


FIGURE 1

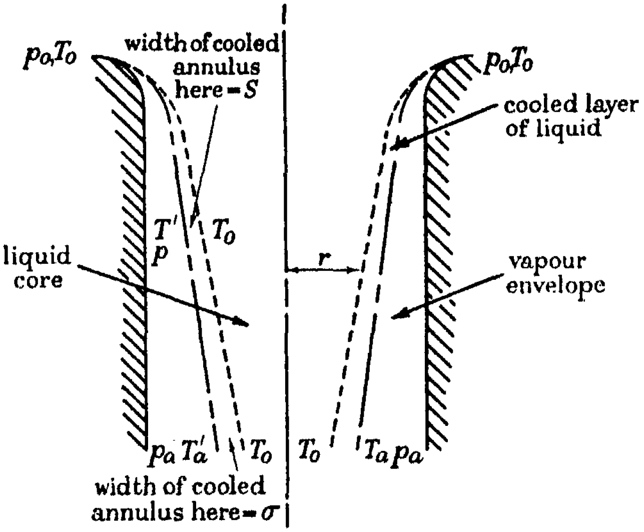


FIGURE 2

However, the difficulty may be circumvented as follows. Making use of the fact that the time is very small and that therefore the gradient near $r = a$ in figure 1*b* is very steep, the approximation is made that all the cooling is confined to a thin laminar layer across which the temperature is linear, i.e. the distribution shown in figure 1*c* as an approximation equivalent to figure 1*b*.

The thickness S of the cooled layer must vary from zero at the entrance to σ at the exit. Moreover, since the pressure must fall from p_0 at entrance to p_a at exit, there will also be an axial variation of surface temperature T'. Hence the complete axial conditions are as represented in figure 2. For the present attention will be confined to the temperature in the liquid only. The temperature in the vapour envelope demands special consideration and will be dealt with later. Meanwhile it may be noted only that at a point where the pressure is p, the surface temperature of the liquid must be higher than the saturation temperature T corresponding to p, to permit evaporation to occur. It is for this reason that surface temperature has been denoted by stressed symbols. The variation of pressure p from p_0 to p_a applies of course to both liquid and vapour, and since frictional effects are neglected, p is assumed to be uniform across a section. The conditions represented by figure 2 can now be summarized:

(1) At any intermediate point between entrance and exit:
   Pressure $= p$.
   Thickness of cooled annulus $= S$.
   Temperature of liquid surface $= T'$.
   Temperature gradient at surface $= (T_0 - T')/S$.
   Radius of liquid core $= r$.

(2) At entrance:
   Pressure $= p_0$.
   Thickness of cooled annulus $= 0$.
   Temperature of liquid surface $= T_0$.
   Temperature gradient at surface $= (T_0 - T_0)/0 =$ indeterminate.
   Radius of liquid core = nozzle radius.

(3) At exit:
   Pressure $= p_a$.
   Thickness of cooled annulus $= \sigma$.
   Temperature of liquid surface $= T'_a$.
   Temperature gradient at surface $= (T_0 - T'_a)/\sigma$.
   Radius of liquid core $= r_a$.

(4) The radius of the nozzle itself will be denoted by r_n. In practice it has to vary at the inlet, to make up for the vena contracta effect. But the alteration which the inlet makes to the mean value of geometrical radius along the nozzle is small and r_n can be treated as a constant.

Thus at the exit the quantity of liquid flowing in an annulus of thickness σ has been cooled to a mean temperature of $\frac{1}{2}(T_0 + T'_a)$. The quantity emerging per second is $2\pi r_a v_a \sigma / V_l$, where v_a is the velocity at exit, and so the amount of heat which is lost by the liquid per second is

$$H = \frac{2\pi r_a \sigma v_a C_l}{V_l}\left[T_0 - \frac{T_0 - T'_a}{2}\right]$$

$$= \frac{\pi r_a \sigma v_a C_l}{V_l}(T_0 - T'_a). \tag{15}$$

But this rate of heat loss is due to the conduction across the cooled layer, integrated for the whole surface from entrance to exit. Hence we also have

$$H = 2\pi k \int_0^{l_n} \frac{r(T_0 - T')\,dl}{S}.$$

Now the rate of fall of pressure along l is not known and therefore a law cannot be given for the variation of r, T' and S with l. An approximation has to be made by considering probable averages.

As $T_0 - T'$ goes from 0 to $T_0 - T'_a$ let it be assumed that its average value $T_0 - \bar{T}$ is of order $\frac{1}{2}(T_0 - T'_a)$. Similarly for $\bar{S}$ take $\frac{1}{2}\sigma$, and for $\bar{r}$ take $\frac{1}{2}(r_n + r_a)$. Thus

$$H = \frac{\pi k l_n (r_n + r_a)(T_0 - T'_a)}{\sigma}. \tag{16}$$

Equating (16) to (15),
$$\sigma^2 = \frac{k l_n V_l}{v_a C_l}\left(1 + \frac{r_n}{r_a}\right). \tag{17}$$

Because of the low heat transfer through the nozzle wall, it can be shown that heat conduction radially across the vapour envelope is small compared with the heat required for the observed evaporations. Accordingly the whole of the rate of heat transfer H will be equated to the rate of vapour formation.* If W is the total weight of fluid passing through the nozzle per second and if a proportion q is evaporated, then

$$H = qWL. \tag{18}$$

But
$$W = \frac{\pi r_n^2 v_a}{V_l + qV_a},$$

since r_n is the exit radius of the nozzle,

$$H = \frac{qL\pi r_n^2 v_a}{V_l + qV_a}. \tag{19}$$

Now we equate (19) and (15) and obtain

$$\sigma C_l (T_0 - T'_a) = \left(\frac{qLV_l}{V_l + qV_a}\right)\frac{r_n^2}{r_a}. \tag{20}$$

We now substitute for σ from (20) into (17) and obtain

$$\frac{q^2 L^2 V_l^2 r_n^4}{(V_l + qV_a)^2 r_a^2} = \frac{k l_n V_l C_l}{v_a}\left(1 + \frac{r_n}{r_a}\right)(T_0 - T'_a)^2. \tag{21}$$

But since the velocity of core and envelope is assumed uniform, and q is small

$$\frac{r_a^2}{V_l} = \frac{r_n^2}{(V_l + qV_a)}. \tag{22}$$

---

* The difference between heat transfer rate in the liquid and in the gas is equal to the heat used in evaporation per second, i.e. $H_1 - H_g = qWL$. But H_g is small compared with H_1 so that equation (18) can be stated. Because of the low condensation coefficient combined with an accommodation coefficient of unity, the rate of vapour molecules striking the surface is many times more than the rate of net evaporation. Since every molecule striking the surface attains temperature equilibrium, the heat taken away for any supposed temperature difference ΔT between surface and vapour by vapour striking the surface is very much more than is needed to raise the actual net quantity of vapour flowing through the space by ΔT. A balance will be struck depending on the rate of heat loss through the metal wall of the nozzle. If this is small, the vapour envelope will be heated very nearly to T'. Because of the thickness of the metal and the low temperature difference between the vapour inside and the outside vapour atmosphere, the rate of heat loss through the nozzle walls in the experiments is small compared with the rate of heat supply corresponding to the observed evaporations. Ultimately therefore, the possibility of neglecting heat transfer across the vapour envelope depends upon the radial transfer across the nozzle wall.

Using (22) and (9) in (21) gives eventually

$$\frac{q^2(1+mq)^{\frac{1}{2}}}{(1+qV_a/V_l)+(1+qV_a/V_l)^{\frac{1}{2}}}=\frac{kl_nV_lC_l(T_0-T'_a)^2}{L^2r_n^2v_c}. \tag{23}$$

Hence equation (23) is a quadratic in $(T_0-T'_a)$, and if q is measured experimentally, and p_0, T_0 and p_a are known, it can be solved and so T'_a is obtained.

It is easy to determine q experimentally by measuring in a tank the weight W of liquid discharged per second through a nozzle from a vessel at pressure p_0 and temperature T_0:

$$\text{for}\quad W=\frac{\pi r_n^2v_c(1+mq)^{\frac{1}{2}}}{V_l+qV_a} \tag{24}$$

$$\text{and}\quad v_c=\sqrt{\{2V_l(p_0-p_a)\}}.$$

V_a and T_a are respectively the saturation vapour volume and temperature corresponding to p_a. Hence if p_a is known, q can be calculated from a knowledge of W.

It is seen therefore that a knowledge of p_a is fundamental both to the determination of q and to the subsequent solution of (23) for T'_a which also involves v_c and V_a. It would normally be assumed that p_a, which is the pressure at the end of the nozzle, would be equal to the pressure in the space outside the nozzle into which the latter discharges. But it will now be proved that this is not always true, and that the nozzle exhibits a critical pressure similar to that of a nozzle discharging a gas. The special interest of the saturated liquid case is that it occurs at low velocities, of the order 20 m./sec., while with a gas the velocity at which the critical pressure occurs is of the order 400 m./sec.

For the remainder of the paper attention will be directed towards the critical pressure in the nozzle, since that must be known before q and therefore T'_a and W can be obtained.

4. CRITICAL PRESSURE

From equation (24) values of W can readily be calculated for a series of values of q for a given p_0 and an assumed p_a. Values of $(T_0-T'_a)$ can similarly be calculated from equation (23) for the same series of values of q for the same given p_0 and the same assumed p_a. Accordingly, a series of corresponding values of W and of $T_0-T'_a$ is obtained and W may be plotted against $T_0-T'_a$. The same process can be repeated for the same given p_0 but for a different value of p_a, and another series of corresponding W and $T_0-T'_a$ obtained. In figure 3 are shown a set of typical curves of W against $(T_0-T'_a)$ found for water by calculating for a given p_0 and the different values of p_a. To each value of p_a a particular curve corresponds.

Now if T_a is the saturation temperature corresponding to the assumed nozzle end-pressure p_a then T_0-T_a is known and its value can be spotted on the curve for the value p_a. In figure 3 the values of T_0-T_a corresponding to the various values of p_a are indicated by circles which have been joined up by the dotted curve.

Now it is known that T'_a must be greater than T_a, since the surface temperature cannot fall below the saturation value corresponding to the pressure p_a. Hence $T_0-T'_a$ is necessarily less than T_0-T_a. It follows that on the lines in figure 3 the only possible values of W for a particular p_a are those which lie to the left of the dotted curve. The values indicated by the dotted curve are the values of W which would be obtained if the surface temperature could fall so low as to attain the value T_a.

The dotted curve exhibits the very interesting property of having a maximum discharge. The values of W indicated by it first increase as p_a is reduced and then diminish. It is readily appreciated from the nature of the graphs that the points corresponding to the true values of W obtained at the T'_a corresponding to each T_a must show a similar maximum somewhat higher in amount and lying to the left of the maximum of the dotted curve. Owing to the slopes of the individual curves for W against $T_0 - T'_a$, however, it will be apparent that the value of $T_0 - T'_a$ for which the true maximum discharge occurs will correspond to a value of $T_0 - T_a$ not very different from that corresponding to the pressure at which the dotted curve shows a maximum. In other words the actual discharge from the nozzle although it occurs with a surface temperature T'_a may be expected to have a maximum at a nozzle end pressure very close to, if not the same as, that corresponding to the maximum of the dotted curve. The value of the maximum flow will be rather greater than indicated by the maximum of the dotted curve.

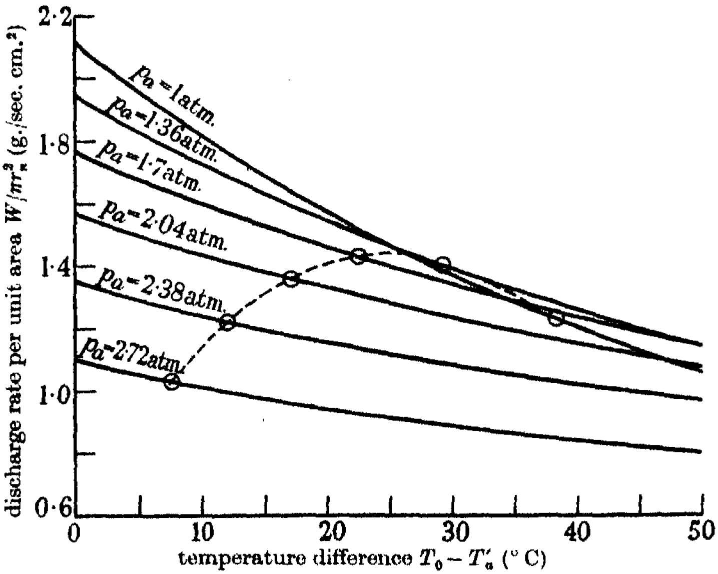


FIGURE 3. Graph of relation between discharge rate and $T_0 - T'_a$ as given by equations (23) and (24). Nozzle length 1·35 cm., diameter 0·476 cm., supply pressure $p_0 = 3{\cdot}4$ atm. Similar graphs have been calculated for each nozzle and each experimental supply pressure. ---- curve connecting the circles which denote $T_0 - T_a = T_0 - T'_a$, where T_a is the saturation temperature corresponding to the particular p_a.

The theory which has led to equations (23) and (24) therefore leads to the conclusion that as the nozzle end-pressure is progressively reduced the calculated discharge first increases and then begins to diminish after passing through a maximum. The value of the back pressure corresponding to the maximum discharge can be determined from the maximum of the dotted curve but the actual value of the maximum discharge is so far uncertain. The important thing at present, however, is not its amount but the fact of its existence. The occurrence of a maximum in a graph of calculated discharge against nozzle end-pressure in the flow of any fluid implies that dW/dp_a is zero. Now it is readily proved that this condition can be satisfied only if the velocity of efflux is equal to the velocity of pressure propagation in the fluid. For the discharge of any fluid, the weight flowing per unit area per unit time is v/V and when the derivative of this with respect to pressure is equated to zero, leads to the condition

$$\frac{dv}{dp} = \frac{v}{V}\frac{dV}{dp}. \tag{25}$$

The general equation for flow without friction gives $vdv = -Vdp$. Hence

$$\frac{dv}{dp} = -\frac{V}{v}.$$

Substitution in (25) gives

$$v^2 = -V^2\frac{dp}{dV}. \tag{26}$$

But the expression on the right-hand side is the general expression for the square of the velocity of pressure propagation in a fluid. In that expression the conditions of the propagation, as for example whether it is adiabatic or isothermal, are defined by the derivative dp/dV and the expression is valid for any condition so defined. Hence whatever the conditions under which the flow of a fluid is assumed to take place, if the calculations indicate a maximum discharge, then it must correspond to the attainment of a velocity of efflux equal to the velocity of pressure propagation in the particular conditions of flow.

It follows that in practice a further reduction of pressure in the space beyond the nozzle will not be experienced in the nozzle and so the nozzle end-pressure should remain at the value corresponding to the maximum discharge and the discharge should remain at that maximum no matter how much the outside pressure is further reduced.

Hence the above argument shows that a nozzle through which a saturated liquid is discharging should exhibit a critical pressure similar to that known in discharge of a gas. However, it has rather different properties. In the case of a gas the critical pressure as calculated for adiabatic flow is a function of the fluid alone. When the foregoing theory is applied to a saturated liquid, however, a different graph of the type shown in figure 3 has to be worked out for each length of nozzle and the critical pressure at which the maximum discharge occurs is different for different lengths of nozzle. The expressions are too complicated to permit of determining a critical condition by differentiation to get dp/dV. Hence the need for graphical calculation of the maximum. In figure 4 a graph is shown of the critical pressure calculated for saturated water flowing through nozzles of varying length and of 0·476 and 1·27 cm. diameter. The curves shown are for a supply pressure of 3·4 atm. with a corresponding saturation temperature of 138·2° C. It is seen that the critical pressure calculated from the theory is dependent not only on the length of the nozzle but on its diameter.

Now the theory has been based upon the conception of a core of liquid and an envelope of vapour. For liquids the velocity of pressure propagation is normally in the region of 1500 m./sec. and for gases it is in the region of 400 m./sec. The velocities of efflux of the saturated liquid which correspond to the calculated values of critical pressure can also be determined from the foregoing theory and it is found that they are very much lower than would normally be considered reasonable for velocities of pressure propagation. Thus, for example, the velocities corresponding to the critical pressures in figure 4 have been plotted in figure 5. They are of the order of only 20 m./sec. It is therefore of considerable interest to investigate experimentally whether the phenomenon of critical pressure does in fact occur under these conditions and to see whether it shows the predicted dependence upon length and upon diameter.

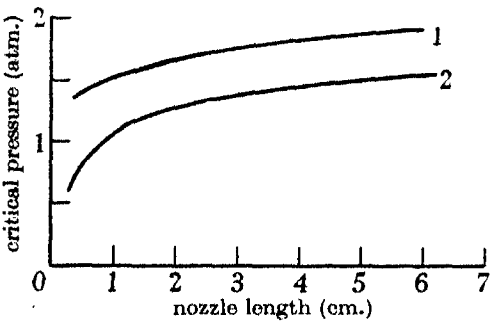


FIGURE 4. Variation of critical pressure with nozzle length. Typical results calculated for water initially saturated at a supply pressure of 3·4 atm. Curve 1, diam. 0·476 cm.; curve 2, diam. 1·27 cm.

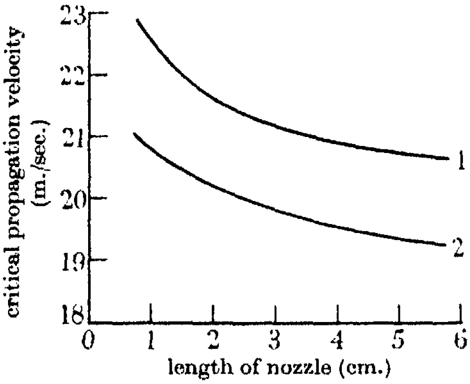


FIGURE 5. Critical propagation velocities (calculated to correspond with the critical pressures in figure 4). Curve 1, diam. 1·27 cm.; curve 2, diam. 0·476 cm.

5. EXPERIMENTAL RESULTS

The experimental apparatus illustrated diagrammatically in figure 6 was used. The pressure vessel was of cast iron and in it steam and water were mixed to obtain water approximately at saturation temperature. The steam supply was taken from the 350 lb./sq.in. main in the laboratory while water was supplied to the experimental vessel by a by-pass from a centrifugal pump line. The main discharge of the centrifugal pump was maintained at the pressure of 6·45 atm. and the use of the by-pass permitted fine control of the water supply to the experimental vessel for operating it at any pressure up to 3·5 atm., the maximum value used. The discharge from the nozzle was measured by collecting it, after condensation of any steam present, in a weighing tank, and measuring with a calibrated stop watch the time for 25·4 kg. (½ cwt.) to collect. Each result was the mean of 5 determinations. Experimental scatter was found to be about 5 %.

Experimental nozzles were made of monel to avoid corrosion. All were designed with an entrance as shown in figure 7 and the lengths used were as follows, in centimetres—0·715, 1·35, 1·98, 2·62, 4·52 and 5·8. As shown in figure 6 the first set of experiments was so arranged that the pressure outside the nozzle to which it was discharging was always one atmosphere. Experiments were made at various supply pressures but from the point of view of the present study of critical pressure only those are of interest in which the predicted critical values are higher than atmosphere. As shown in figure 4 this is the case for all the lengths when the supply pressure is 3·4 atm. and the experimental results will therefore be described for this value.

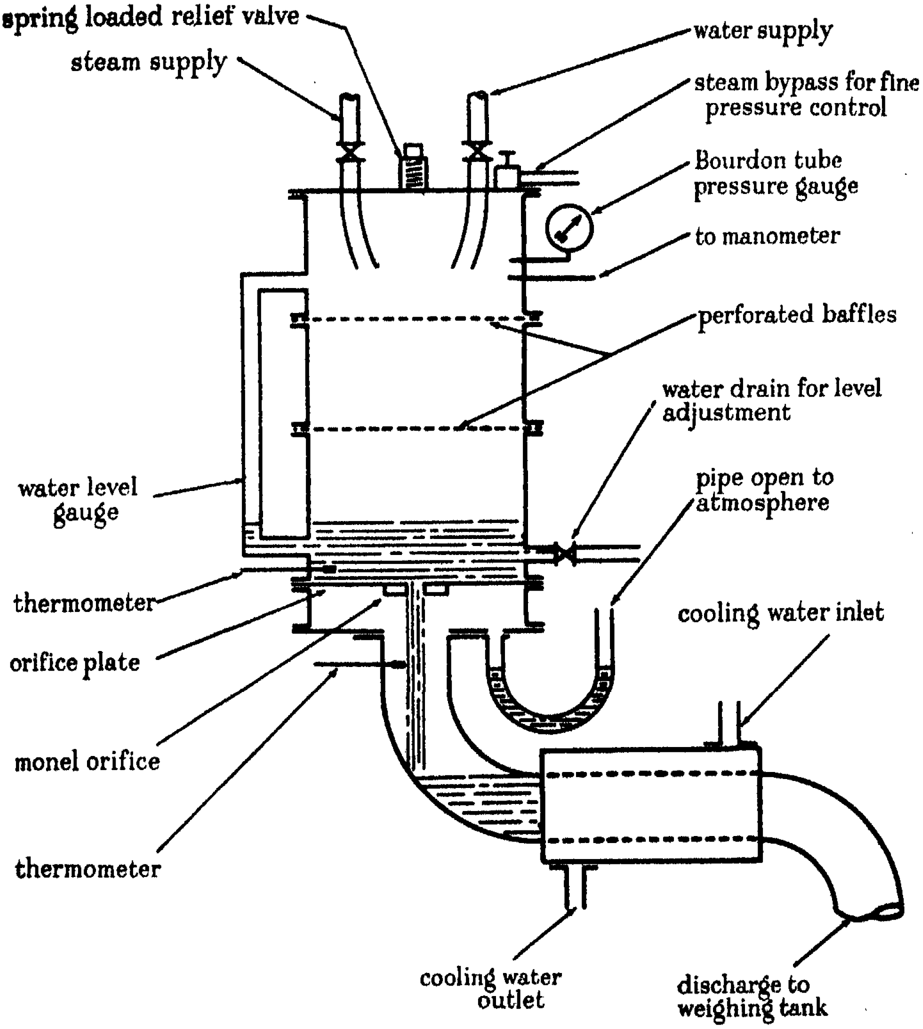


FIGURE 6. Diagrammatic arrangement of experimental apparatus.

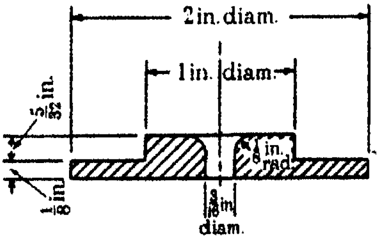


FIGURE 7. Detail of rounded entrance nozzle.

The experimental discharge obtained with a supply of saturated water at 3·4 atm. for the various lengths is shown in figure 8, and compared with theoretical values calculated from (24). It is immediately evident that the discharge is very much affected by the length of the nozzle and that general agreement is good. Hence if a critical pressure does exist it cannot be independent of nozzle length since if it were so the discharge would remain unmodified by length. The next step was to measure the pressure inside one of the nozzles but near its exit to see whether the pressure at its end was in fact higher than atmosphere. Accordingly a small radial hole was drilled through the wall of the longest nozzle of 5·8 cm. and connected to a manometer. When the fluid was now allowed to discharge as before the end pressure in the nozzle as measured by the manometer was found to be 2·39 atm. There was therefore no doubt about the existence of a critical value.

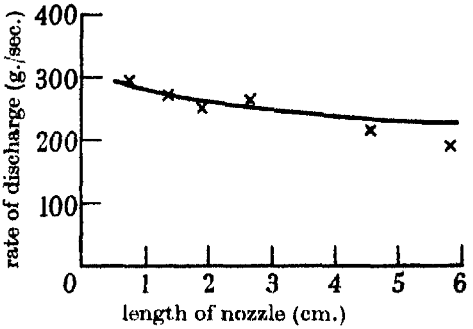


FIGURE 8. × experimental results for nozzle diam. 0·476 cm. at supply pressure 3·4 atm. —— calculated discharge for varying lengths.

Unfortunately it was found that the discharge was no longer the same as before the pressure measurement hole had been drilled. Instead of 189 g./sec. it was found to be only 156 g./sec. A new nozzle was made of the same length and diameter but again without a pressure measurement hole and the previous result of 189 g./sec. was found to be repeated. The experiment with the pressure measurement hole was repeated also and again gave 156 g./sec. It has therefore to be accepted that the existence of the pressure measurement hole disturbs the metastable jet in some way so as to reduce its stability and cause more rapid formation of vapour. Accordingly the value of 2·39 atm. read for the critical pressure cannot be taken as a correct measurement of the critical pressure corresponding to the discharge of 189 g./sec. in the original nozzle. Nevertheless the point of importance is that the existence of a critical pressure has been proved for conditions which are not very different so far as velocity of efflux is concerned from those in the nozzles without pressure measurement points. For it is easy to calculate from the theory, knowing the nozzle end pressure as 2·39 atm., the proportion of vapour and the resultant velocity which could give the observed discharge of 156 g./sec. The result is found to be 17·7 m./sec.

It may therefore be considered established that a critical pressure can exist under the conditions of discharge of saturated water and the critical pressures can be estimated which did exist in the experiments even although it is not possible to measure them without upsetting the stability. Such an estimate can be made by considering a set of figures of the type of figure 3 for various lengths and plotting the indicated maximum discharge against the indicated critical pressure. It has already been mentioned that the actual discharge through the nozzle at the critical pressure should be rather higher than these values but the difference will not be great. Accordingly it may be assumed that the relation so established between discharge quantity and critical pressure will be approximately true and will hold generally for the given diameter. Support for this assumption is obtained in that the critical pressure corresponding to a discharged quantity of 156 g./sec. is predicted as 2·34 atm. which compares favourably with the actual measured value of 2·39 atm. The estimated experimental critical pressures can then be compared with the theoretical prediction of figure 4. The following table shows the comparison:

length in cm.	0·715	1·35	1·98	2·62	4·52	5·8
critical pressure: theor.	1·50	1·574	1·645	1·715	1·84	1·95
exp.	1·46	1·545	1·62	1·565	1·94	2·18

It is therefore established that an experimental critical pressure is obtained which follows the predicted trend with length of nozzle.

For nozzles of larger diameter the quantity discharged was so great that it was difficult to make many experiments and to maintain the same accuracy. However, by the same methods as above it was established that for a nozzle of 1·27 cm. diameter and 5·8 cm. length the experimental critical pressure was 1·95 atm. as compared with a theoretical value of 1·55 atm. When it is recalled that for the same length of 0·476 cm. diameter the corresponding values were respectively 2·18 and 1·95, it will be seen that the experimental critical pressure does follow the prediction that it should diminish with increase in the nozzle diameter.

The critical pressure must be similarly considered for any given supply pressure. A nozzle of given length and diameter will have a critical pressure dependent upon the supply pressure. The critical values can be calculated by the same procedure as outlined above. Now in some experiments it may be that the pressure of the space into which the nozzle discharges is higher than the critical value and when that is so the velocity of pressure propagation will not have been attained and the external pressure will be effective within the end of the nozzle. When the external pressure is below the critical value it will not, however, be effective within the nozzle. In either case the value of the pressure within the end of the nozzle will be whichever is the higher. The quantity of fluid which is discharged in any experiment should be greater than the quantity corresponding to the nozzle end-pressure as read on the appropriate dotted curve of the type shown in figure 3. When the external pressure is below the critical value then the delivered quantity should be slightly greater than the maximum of the dotted curve, while if the external pressure is greater than the critical value the discharged quantity should be greater than the quantity indicated by the dotted curve at the value of $T_0 - T_a$ corresponding to the external pressure. The reason for anticipating an excess over these values may be restated here for convenience. It is due to the fact that the surface temperature of the liquid core has to be at a value of T'_a higher than T_a, in order that evaporation should take place. Hence if experiments are made to determine the quantity discharged and if it is in fact found to be higher than the values indicated by the dotted curve for the appropriate nozzle end-pressure, then the actual value of T'_a may be calculated from the theory by solving equations (23) and (24).

6. PRESSURE PROPAGATION IN A SATURATED LIQUID

Before concluding this paper, it will be of interest to give some general discussion of the theoretical peculiarities of pressure propagation in a saturated liquid, in view of the above experimental evidence for low velocities of propagation. It may be noted first that the fundamental physical reason for low velocities is in the change of phase from liquid to vapour, which causes unusually low values of dp/dV. Then the very important point appears that if one starts with the homogeneous saturated liquid phase, vapour cannot form for positive pressure increments but may form for negative amplitudes. Accordingly, considering the passage of a sound wave, the value of dp/dV may be very much less for the negative than for the positive parts of the cycle. Hence a liquid phase at saturation temperature may have a rectifying property for sound waves, since for positive amplitudes the propagation velocity may be expected to be in the usual liquid region of 1500 m./sec., but may be much less for negative amplitudes. The extent to which this rectifying effect does actually occur will depend upon the extent to which vapour does in fact form. As we have seen in the introduction, and as the experiments with different length of nozzle have proved, that will depend upon the extent of free surface available, and upon the time. Taking the latter point first, the rectification may be expected to depend upon the frequency. Low frequency cycles may permit considerable vapour formation during the diminishing portion of the negative half and equal condensation during the increasing portion of the negative half, giving very low values of dp/dV for that half. But very high frequencies may be too rapid to give dp/dV appreciably different in the negative from the positive side. Accordingly ultrasonic waves may not be at all rectified in saturated liquids and very steep negative wave fronts may be transmitted. Experimental evidence for the latter is available in the propagation of rarefaction waves following underwater explosions. But the extent of free surface must be at least as important as the frequency, because as was suggested in the introduction and proved by the nozzle experiments, the available surface governs the rate of vapour formation. In very turbulent or eddying water, or water with gas dissolved in it, the formation of bubbles to provide surface may be easy. In other cases it may be very difficult. When the difficulty of forming bubbles is prohibitive, the governing surface feature will be the presence of solid particles, sand, grit, etc., dispersed through the fluid. Finally while both frequency and available surface will be important in limiting the propagation of negative parts of pressure waves, the amplitude of the negative part must also be of significance. Mention may also be made of the case where to a cold liquid is applied a pressure wave of amplitude so great that at its lowest value the pressure falls below the saturation point corresponding to the temperature. This case is of frequent practical occurrence, as in the case of ship propellers, pumps, etc. The results described in the paper show the importance of considering the possibility of bubbles or dispersed surface, the frequency, and the amplitude, in any problem involving the transmission of negative pressure through a liquid.

The work formed part of a general investigation of condensation and evaporation in the research programme of Messrs G. and J. Wier, Ltd., to whom I am indebted for permission to publish.

References

Benjamin, M. W. & Miller, J. G. 1941 *Trans. Amer. Soc. Mech. Engrs.* **63**, 419.
Bottomley, W. T. 1937 *Trans. N.-E. Cst. Instn. Engrs. Shipb.* **53**, 65.
Silver, R. S. & Mitchell, J. A. 1945 *Trans. N.-E. Cst. Instn. Engrs. Shipb.* **62**, 51.

The detonation of solid explosives: the equilibrium conditions in the detonation wave-front and the adiabatic expansion of the products of detonation

By H. Jones and A. R. Miller*

Mathematics Department, Imperial College of Science and Technology, London

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 20 *November* 1947)

Assuming chemical and thermal equilibrium to be maintained in the detonation wave-front, and using the equation of state in the form of the virial expansion, the velocity of detonation has been determined as a function of the loading density. In the absence of data at sufficiently high pressures and temperatures for the products of detonation of T.N.T., it has been assumed that the virial coefficients are constant and their values have been determined to give agreement with the measured values of the detonation velocity for loading densities less than 1·5 g.cm.$^{-3}$. The pressure-volume-temperature relation in the detonation wave-front can then be determined. The pressure in the detonation wave-front is found to be of the order of 2×10^{11} dyne cm.$^{-2}$ for a loading density of 1·5 g.cm.$^{-3}$, compared with the value of $9{\cdot}4\times10^{10}$ dyne cm.$^{-2}$ given in the earlier work of other authors using the co-volume method. With the equation of state adopted in this paper it is found that at a high loading density only negligibly small amounts of hydrogen and carbon monoxide are present in the detonation wave-front, a fact which facilitates the calculation of the adiabatic relations in this case. It is shown (part B) that these gases do, however, develop rapidly during the initial stages of the adiabatic expansion.

The calculation of the equilibrium conditions in the detonation wave-front with the adopted equation of state (part A) determines the initial conditions for the calculation of the adiabatic relations for a high loading density. The chemical composition of the gases during the adiabatic expansion and the external work done during it have been calculated for a loading density of 1·5 g.cm.$^{-3}$ (part B). It is shown that the large amount of chemical energy

---

* Now at the Royal Society Mond Laboratory, Cambridge.

released in the early stages of the expansion is to be correlated with the high effective value of the exponent in the adiabatic in this region, and this is due to the dominant role of the repulsive forces between the molecules of the tightly compressed gases during the early stages of the expansion. The same effect is also observed in the case of a low loading density (part C).

The difference with regard to the amounts of hydrogen and carbon monoxide present in the detonation wave-front for a low loading density complicates the solution of the equations in this case. In part C it is shown how this can be done for a loading density of 1·0 g.cm.$^{-3}$, and the detonation velocity and the pressure, density and temperature in the wave-front have been determined using the same equation of state as in parts A and B. The adiabatic pressure-volume relation for the expansion of the products of detonation, and the chemical composition during, and up to the end of, the adiabatic expansion have also been determined. Compared with the results for a high loading density, there is considerably more carbon monoxide and less carbon dioxide and a substantial rise in the total number of moles of gas produced per mole of explosive. The ratio is found to be sensitive to the pressure in the detonation wave-front, from which, by comparison with the observed values of this ratio, independent evidence is obtained for the detonation pressure calculated in part A. The chemical energy released per gram of explosive is less for a loading density of 1·0 g.cm.$^{-3}$ than for a loading density of 1·5 g.cm.$^{-3}$, and the external work done is also less in the former than in the latter case.

The amounts of ammonia and of hydrocyanic acid in chemical equilibrium with the other gases are determined and they are found to be negligibly small. It is concluded that these gases, observed in experiments, are probably formed by catalytic action with the bomb fragments during the cooling period after the adiabatic expansion has been completed. The calculations have been compared with the available experimental data and are in reasonable agreement with it. An explanation is suggested for the observed difference in the composition of the gaseous products of the detonation of T.N.T., initiated at a given loading density, with detonators of different power.

1. INTRODUCTION

1·1. For a solid explosive the pressure, temperature and molecular volume in the detonation wave-front have values which lie far outside the range which has been studied in other connexions, so that little guidance is available for the purpose of setting up the equation of state appropriate to the conditions in the detonation wave-front. Mainly for this reason, but also because the products of the detonation of a solid explosive are more complicated than those of the detonation of a gas, the study of the process of the detonation of solids presents greater difficulties than of that in gases.

1·2. Many investigations of the properties of explosives have made use of a form of van der Waals's equation (Cranz 1926; Crow & Grimshaw 1930), which is obtained from that for a perfect gas by diminishing the specific volume by a certain constant amount referred to as the co-volume. This has been used widely for the interpretation of experiments in which the pressure is measured after an explosion in a closed bomb. The value of the co-volume obtained in this way is much less than that determined under ordinary conditions, and the discrepancy is generally too great to be accounted for by the high temperature of the explosion. Although this equation of state fits the experimental results obtained in closed bomb experiments approximately (Cranz 1926, p. 40), it is entirely inadequate for the study of the detonation process. For, immediately behind the detonation wave-front the density is greater than that of the unexploded material, and the molecular volume is actually less than the excluded volume which it would be necessary to assume in applying this equation of state. Furthermore, it requires that the internal energy shall be a function of the temperature only, but under the high compression which exists in the detonation wave it is clear that the repulsive forces between the tightly packed molecules will make a significant contribution to the energy. It is, therefore, clear that a theory of detonation with this equation of state as basis can only be built up by fixing the value of the co-volume arbitrarily so as to give the observed rate of detonation (cf. Langweiler 1938); co-volumes so determined can have little physical significance.

1·3. In this paper the detonation velocity is determined as a function of loading density by assuming chemical and thermal equilibrium to be maintained in the wave-front. By comparing this relation with the observed values of the detonation velocity the virial coefficients in the equation of state have been determined. The pressure, temperature and volume in the detonation wave-front have then been determined. The initial conditions calculated in this way can then be used as the starting-point from which to calculate the pressure-volume-temperature relations and the chemical composition during the adiabatic expansion of the material in the detonation wave-front of a solid explosive at different loading densities. Detailed application is made to T.N.T., and the effect of varying the loading density is determined.

1·4. In determining the adiabatic it is assumed that complete chemical and thermal equilibrium exists at each instant during the expansion. The validity of this assumption must be judged by the results to which it leads, but in any case it is a useful assumption to make since it provides a precisely defined limiting case. It should, however, be noted that while it is probable that complete chemical equilibrium exists in the initial stages of the expansion it is certain that in the later stages chemical equilibrium is not maintained. The experiments of Robertson & Garner (1923), for example, show that the relative amounts of the product gases when they have cooled to room temperature still correspond to an equilibrium temperature of about 1800° K. It would appear at first sight that the calculation of the adiabatic relations is faced with an almost insuperable arbitrariness in that, without a detailed knowledge of the reaction velocities of the various reactions taking place, it is not possible to determine with certainty the pressure and temperature under which the chemical equilibrium is to be considered to be frozen out. However, when the composition is determined during adiabatic expansion assuming chemical equilibrium it is found that the amount of free carbon present decreases rapidly at first as the volume increases, and then, passing through a minimum, it increases slowly and steadily. At about the same stage as there is a minimum of free carbon there is a maximum of carbon monoxide and of the total amount of gas present. Further, it is found that, in this region of the adiabatic expansion, the composition changes only very slowly in relation to the volume or the pressure. The reaction velocities are known to decrease rapidly with decreasing pressure and temperature; accordingly, in the present calculations, it is assumed that if chemical equilibrium is maintained up to the stationary values of the composition, the reaction velocities will then have fallen to values negligible compared with the speed of the expansion before the composition has changed appreciably from that at the stationary point. More precisely, it is assumed that chemical equilibrium exists up to the stage at which there is a minimum of free carbon and that thereafter the composition is fixed.

PART A. EQUATIONS DETERMINING THE CONDITIONS IN THE DETONATION WAVE-FRONT FOR SOLID EXPLOSIVES

2. ENERGY EQUATION

The equations of continuity of mass and momentum across the wave-front can be written

$$U = (p\rho')^{\frac{1}{2}}/\rho_0, \tag{1}$$

$$u = (p/\rho')^{\frac{1}{2}}, \tag{2}$$

with

$$\frac{1}{\rho'} = \frac{1}{\rho_0} - \frac{1}{\rho}, \tag{3}$$

where U is the velocity of the detonation wave-front, u is that of the material in the wave-front; ρ_0 is the density of the unexploded material, ρ is that of the material in the wave-front; p is the pressure in the wave-front, and the pressure of the unexploded material is taken as zero since atmospheric pressure is completely negligible compared with p.

As the wave-front advances through the solid explosive, solid material passes into it and acquires a velocity u, and is subject to work done by the pressure difference p. At the same time chemical energy, H per mole of explosive, is released in the chemical reactions which take place in the wave-front. Thus if the molecular weight of the explosive be denoted by M, the total internal energy per gram of material in the detonation wave-front is given by

$$\frac{H}{M} + \frac{pu}{\rho_0 U} - \tfrac{1}{2}u^2,$$

and, using equations (1), (2) and (3), this reduces to

$$E - H = \tfrac{1}{2}p(v_0 - v), \tag{4}$$

where v_0 and v are the molecular volumes and E is the internal energy of the exploded material per mole. Equation (4) is generally referred to as the Hugoniot equation. The energy of the material in the wave-front is a function of the molecular volume and the temperature or, when the equation of state is known, it may be expressed as a function of the pressure and the volume. The energy of the chemical reaction can also be expressed as a function of the pressure and the volume. Thus, the Hugoniot equation expresses a relation between the pressure and the volume in the detonation wave-front. To determine them a second relation is required. This is provided by the Jouguet (1906) condition which is usually expressed by the equation

$$\left(\frac{dp}{dv}\right)_h = -\frac{p}{v_0 - v}, \tag{5}$$

in which $(dp/dv)_h$ denotes the gradient of the line which represents the Hugoniot equation in the p-v diagram.

3. EQUATION OF STATE

It was explained in the introduction that there is very little information available to indicate the correct form of the equation of state to be used in calculations dealing with the detonation wave. First, it is to be noted that there is no virtue in using any of the usual complicated empirical equations of state. For the complicated form of these equations is designed to cover the critical phenomena which are controlled by the weak molecular attractions, while in the range of pressure and temperature which exist in the detonation wave these weak molecular attractions are negligible compared with the mean molecular energies. The best that can be done therefore is to choose the simplest type of equation which is yet sufficiently general to include the effects of molecular repulsion on the internal energy. In this paper the virial equation in powers of the pressure, terminated at the term in p^3, is used, so

$$p\frac{v'}{N'} = RT + bp + cp^2 + dp^3, \tag{6}$$

where v' is the volume and N' the number of moles at temperature T and pressure p of the gaseous products of the detonation of 1 mole of explosive. To assume a constant co-volume, as is done by Schmidt (1935, 1936) and by Langweiler (1938), corresponds to terminating the expansion at the first power. The virial coefficients b, c and d are, of course, functions of the temperature and of the mole fractions of the constituents of the gas. In the absence of any specific knowledge of the nature of these functions in the range of pressure and temperature in the wave-front, or for the products of detonation, the drastic assumption of regarding them as constants is made. On the basis of this assumption a relation can be determined between the detonation velocity and the loading density and is given by equations (1) and (2) with the values of p and v determined from equations (26), (27) and (28) below. The values of the virial coefficients are then determined by comparing the calculated relation between detonation velocity and loading density with the observed values, for loading densities less than 1·5 g.cm.$^{-3}$. It is probably best to state immediately the values which are obtained for the virial coefficients when this procedure is carried out. The result is

$$b = 25{\cdot}4, \quad c = -0{\cdot}104, \quad d = 2{\cdot}33\times 10^{-4}, \tag{7}$$

when 1000 kg.cm.$^{-2}$ is taken as the unit of pressure.

4. HEAT ENERGY OF THE PRODUCTS OF DETONATION

4·1. Let there be N_A moles of species A (all A) derived from 1 mole of explosive, and N moles in all, so that

$$N = \sum_A N_A, \tag{8·1}$$

and the total energy per mole of explosive is given by

$$E = \sum_A N_A E_A + N'E_1, \tag{8·2}$$

where N' is the total number of moles of gaseous products, that is, $N' = N - N_s$, where N_s is the number of moles in the solid state. The E_A are defined to be functions of the temperature only. If A denotes a molecular species which is in the gaseous phase in the detonation products, then E_A denotes the energy of this gas at pressures low enough to ensure that the energy depends only on the temperature. If A denotes a species which is in the solid phase in the detonation products, then E_A is taken to be the internal energy at atmospheric pressure. The contribution to the total energy arising from the compression of any solid phase is neglected throughout these calculations. E_1 is the energy per mole of gas arising from the mutual interactions between the gas molecules.

From equations (6) and (7) and the well-known thermodynamical relation

$$\left[\frac{\partial}{\partial p}(E+pv)\right]_T = v - T\left(\frac{\partial v}{\partial T}\right)_p,$$

it follows immediately that

$$E_1 = -\tfrac{1}{2}cp^2 - \tfrac{2}{3}dp^3, \tag{9}$$

since by definition E_1 vanishes at zero pressure.

4·2. We now express the heat energy as a function of the pressure, volume and the number of moles of the different species, taking into account the field of force of the molecules but assuming that it is the same for all molecules. To do this we make use of the relations

$$\left(\frac{\partial G}{\partial N_A}\right)_{T,p,N_B,\ldots} = \mu_A, \quad \left(\frac{\partial G}{\partial p}\right)_{T,N_A,N_B,\ldots} = v,$$

where G is the Gibbs function for the detonation products and μ_A is the partial potential of the species A (all A). The assumption that all the molecules have the same field of force is expressed by the equation

$$\left(\frac{\partial v}{\partial N_A}\right)_{T,p,N_B,\ldots} = \left(\frac{\partial v}{\partial N_B}\right)_{T,p,N_A,\ldots} = \frac{v'}{N'}, \tag{10}$$

where $A, B, \ldots$ denote molecular species in the gaseous phase and N' is the total number of moles of gas in the volume v'. Thus

$$\left(\frac{\partial \mu_A}{\partial p}\right)_{T,N_A,\ldots} = \frac{v'}{N'} \quad \text{(all } A\text{)}. \tag{11}$$

A quantity p^* is now defined by the relation

$$\mu_A = \mu_{A,0} + RT\log(p^*/p), \tag{12}$$

where $\mu_{A,0}$ is the partial potential of the species A in the perfect gas condition, so

$$\left(\frac{\partial \mu_{A,0}}{\partial p}\right)_{T,N_A,\ldots} = \frac{RT}{p}.$$

Thus

$$\left(\frac{\partial \log p^*}{\partial p}\right)_T = \frac{v'}{RTN'}.$$

Integrating this and using the equation of state we get

$$\log\frac{p^*}{p} = \frac{bp + \tfrac{1}{2}cp^2 + \tfrac{1}{3}dp^3}{RT}, \tag{13}$$

in which the constant of integration has been determined by the condition that p^* and p vanish together. For a perfect gas it is known that

$$\mu_{A,0} - RT\log(N_A/N') - RT\log p \tag{14·1}$$

is a function of the temperature only, and therefore for the material of the detonation wave

$$\mu_A - RT\log(N_A/N') - RT\log p^* \tag{14·2}$$

is the same function of the temperature. Now, the equilibrium condition for the reaction

$$A + B \rightleftharpoons AB$$

between perfect gases is

$$\mu_{A,0}dN_A + \mu_{B,0}dN_B + \mu_{AB,0}dN_{AB} = 0.$$

Since the number of atoms is conserved this can be written as

$$\mu_{A,0} + \mu_{B,0} = \mu_{AB,0},$$

whence, for a perfect gas,

$$\log\frac{N_A N_B}{N_{AB}} + \log\frac{p}{N'} = \log K(T), \tag{15}$$

where $K(T)$ is the equilibrium constant for the reaction. Thus for the material of the detonation wave

$$\frac{N_A N_B}{N_{AB}}\frac{p^*}{N'} = K(T) \tag{16}$$

is the required equilibrium condition. Thus the effect of the intermolecular forces is expressed in the equation for chemical equilibrium by replacing the pressure p by the function p^* given by equation (13).

At 3000° K the values given in table 1 are obtained for p^* from equations (6) and (13), assuming the values (to be justified later) for the virial coefficients given by equation (7). It will be noticed that as p increases, and v'/N' decreases, p^* increases very rapidly so that even at high temperatures $N'T/p^*$ is negligibly small for pressures greater than 10^4 atm. This result is of importance in the calculation of the conditions in the detonation wave-front, from equations (19), (20) and (21) below.

TABLE 1. VALUES OF p^* AT 3000° K

p (atm.)	p^*/p	v'/N'
1×10^4	2·88	51·1
5×10^4	98·1	27·5
1×10^5	$2·28\times10^3$	21·6
2×10^5	$2·31\times10^5$	16·5

4·3. From an equation like (16) for each reaction which takes place in the detonation wave-front, the composition of the products of detonation can be calculated and thence the values of E and H can be determined. It is found below that the detonation pressure is of the order of that in the last two lines of table 1, so that in the detonation wave-front p^* is very great and $N'T/p^*$ differs insignificantly from zero. This justifies the procedure of § 5·2 below, and it is for this reason that table 1, which is based on the values eventually determined for the virial coefficients, is included at this stage. This simplifies the calculations which are necessary to determine the conditions in the wave-front greatly. Having determined E and H, the Hugoniot equation and the Jouguet condition can then be used with the equation of state to determine the pressure and temperature in the wave-front and the detonation velocity. This procedure will be illustrated by considering the detonation of solid T.N.T.

5. CHEMICAL EQUILIBRIUM FOR T.N.T.

5·1. It is assumed that no compounds of nitrogen are formed in the wave-front, and, in fact, it will be shown in part C that only negligible amounts of ammonia and hydrocyanic acid are formed even during the adiabatic expansion of the gaseous products of the detonation. With this assumption the decomposition of T.N.T. is represented by the equation

$$\begin{aligned} C_7H_5N_3O_6 = 1{\cdot}5\,N_2 + N_{H_2O}\,H_2O + N_{H_2}\,H_2 + N_{CH_4}\,CH_4 \\ + N_{CO_2}\,CO_2 + N_{CO}\,CO + N_C\,C, \end{aligned} \qquad (17)$$

in which N_A is the number of moles of species A and the carbon is produced as graphite. Consider the reactions

$$CO + H_2O = H_2 + CO_2, \qquad (18{\cdot}1)$$

$$C\,(\beta \text{ graphite}) + 2H_2 = CH_4, \qquad (18{\cdot}2)$$

$$CO + \tfrac{1}{2}O_2 = CO_2, \qquad (18{\cdot}3)$$

$$C\,(\beta \text{ graphite}) + \tfrac{1}{2}O_2 = CO, \qquad (18{\cdot}4)$$

and let $K_1(T)$, $K_2(T)$, $K_3(T)$ and $K_4(T)$ be the equilibrium constants applicable to the perfect gas condition for these four reactions. The values of these constants have been tabulated over a considerable range of temperatures by Lewis & von Elbe (1938, appendix A, table 2). If

$$\log_{10} K(T) + 0{\cdot}4343\Delta E_0^\circ / RT,$$

where ΔE_0° is the energy of reaction at the absolute zero, be plotted against $\log_{10} T$ a straight line is obtained. In this way the tables given by Lewis & von Elbe were extrapolated to higher temperatures.

5·2. The equations which determine the composition are then

$$N_{CH_4} + N_{CO_2} + N_{CO} + N_C = 7, \qquad (19{\cdot}1)$$

$$2N_{H_2O} + 2N_{H_2} + 4N_{CH_4} = 5, \qquad (19{\cdot}2)$$

$$N_{H_2O} + 2N_{CO_2} + N_{CO} = 6, \qquad (19{\cdot}3)$$

$$N_{CO}\,N_{H_2O} = K_1(T)\,N_{H_2}N_{CO_2}, \qquad (20{\cdot}1)$$

$$N_{H_2}^2 = K_2(T)\,(N'/p^*)\,N_{CH_4}, \qquad (20{\cdot}2)$$

$$N_{CO}^2 = \frac{K_3(T)}{K_4(T)}\frac{N'}{p^*}\,N_{CO_2}. \qquad (20{\cdot}3)$$

Eliminating all the molar fractions except that of hydrogen, we obtain the equation

$$(\lambda a_0 + \lambda a_1 f + a_2 f^2)\,(2{\cdot}5 - f - a_3 f^2/\lambda) - 6f^2 = 0, \qquad (21)$$

in which f has been written for N_{H_2}, λ for $N'T/p^*$ and the coefficients a_0, a_1, a_2 and a_3, which are functions of T only, are given by the equations

$$a_0 = \frac{5K_3}{TK_1^2K_4}, \quad a_1 = \frac{K_3(K_1-2)}{TK_1^2K_4},$$

$$a_2 = 1 - \frac{4K_3}{K_1^2K_2K_4}, \quad a_4 = \frac{2T}{K_2}.$$

In the detonation wave-front p^* is very large and the parameter $\lambda = N'T/p^*$ is so small that the composition in the wave-front is given quite accurately by taking $(N'T/p^*)$ to be zero. In this case the total number of moles of gas is constant and equal to 5·75 and it can be verified by using equation (21) that f^2 is of the order of λ and that the composition is given by

$$\left.\begin{array}{c} N_{H_2} = 0, \quad N_{CO} = 0, \quad N_{N_2} = 1{\cdot}5, \quad N_{H_2O} = 2{\cdot}5 - 2\phi(T), \\ N_{CH_4} = \phi(T), \quad N_{CO_2} = 1{\cdot}75 + \phi(T), \end{array}\right\} \tag{22·1}$$

with
$$\phi(T) = \left[\left\{\frac{49}{4} + \frac{240K_3}{K_1^2K_2K_4}\right\}^{\frac{1}{2}} - \left\{\frac{7}{2} + \frac{20K_3}{K_1^2K_2K_4}\right\}\right] \Big/ 4\left\{1 - \frac{4K_3}{K_1^2K_2K_4}\right\}. \tag{22·2}$$

Typical values of $\phi(T)$ are given in table 2.

TABLE 2. VALUES OF $\phi(T)$

T (°K)	1000	2000	3000
$\phi(T)$	0·470	0·603	0·666

This determines the composition in the detonation wave-front at any temperature. We thus get the rather remarkable result that hydrogen and carbon monoxide are present in the detonation wave-front only in negligible amounts. It will be shown in parts B and C below, however, that, due to chemical reaction in the wave-front, both of these gases develop rapidly as the material of the detonation wave-front expands adiabatically. The energies of the gases are known (Lewis & von Elbe 1938, appendix A, table 1), so that the total energy per mole of explosive can be calculated using equations (8). It is found to be very nearly a linear function of the temperature and the energy E kcal./mole T.N.T. can be represented quite accurately by

$$E = 0{\cdot}0875T - 38{\cdot}0 + N'E_1, \tag{23}$$

where E_1 is given by equation (9).

5·3. If the heat of formation of a molecule AB be denoted by ϵ_{AB} and of T.N.T. by $\epsilon_{\text{T.N.T.}}$, all being taken as positive for energetically stable molecules, then the energy liberated by the chemical reaction in the detonation wave-front is given by

$$H = \sum_{AB} N_{AB}\epsilon_{AB} - \epsilon_{\text{T.N.T.}}. \tag{24}$$

For the molecules involved we take the following values (Lewis & von Elbe 1938, appendix A, table 3; Garner & Abernethy 1921, 1930) which all refer to carbon in the form of β-graphite:

$\epsilon_{H_2O} = 57{\cdot}11, \quad \epsilon_{CO_2} = 94{\cdot}16, \quad \epsilon_{CO} = 27{\cdot}40, \quad \epsilon_{CH_4} = 16{\cdot}17, \quad \epsilon_{\text{T.N.T.}} = 15{\cdot}0$ kcal./mole.

Using the values of the N_{AB} given by equations (22) it is found that

$$H = 292{\cdot}57 - 3{\cdot}888\phi(T),$$

so that in the range 1000 to 3000° K, H varies by less than 0·25 %. Accordingly, to calculate the conditions in the detonation wave-front for T.N.T., H has been taken as constant and equal to 290 kcal./mole T.N.T.

5·4. From equations (9), (23) and the equation of state (6), the Hugoniot equation can be written in the form

$$v'\left(\frac{\alpha}{RN'}+\frac{1}{2}\right) = \tfrac{1}{2}v_0'+\frac{\alpha}{RN'}bN'+\left(\frac{\alpha}{RN'}+\frac{1}{2}\right)cpN'+\left(\frac{\alpha}{RN'}-\frac{2}{3}\right)dp^2N'+\frac{H-\beta}{p}, \quad (25)$$

where α is the coefficient of T in the expression for the internal energy. The Jouguet condition leads to the equation

$$v_0' = N'b+2(H-\beta)\Big/p\frac{\alpha}{RN'}-\left(\frac{\alpha}{RN'}+\frac{2}{3}\right)N'dp^2\Big/\frac{\alpha}{RN'}. \quad (26)$$

Thus equations (25) and (26) have to be solved with the equation of state.

From equations (6), (25) and (26), v_0' and v' can be eliminated to yield

$$p^3\left(\frac{\alpha}{RN'}+1\right) = 3(H-\beta)\left(\frac{\alpha}{RN'}+1\right)\Big/N'd-3RT\left(\frac{\alpha}{RN'}+\frac{1}{2}\right)\frac{\alpha}{RN'}\Big/\left(\frac{\alpha}{RN'}+1\right)d. \quad (27)$$

Also, from equations (25) and (26), we obtain

$$v' = v_0'+N'cp-(H-\beta)\Big/p\left(\frac{\alpha}{RN'}+\frac{1}{2}\right)+2\left(\frac{\alpha}{RN'}+\frac{2}{3}\right)N'dp^2\Big/\left(\frac{\alpha}{RN'}+\frac{1}{2}\right). \quad (28)$$

Thus, for a given value of T, equation (27) determines the pressure. Then equation (26) determines v_0' and (28) determines v'. Then we have

$$v_0 = v_0'+N_s v_s, \quad (29)$$

where v_s is the specific volume of the solid phase. Thus for a given loading density, that is, for a given value of v_0, the corresponding values of pressure, volume and temperature in the detonation wave-front can be calculated. The velocity of the detonation wave-front can then be calculated from equations (1) and (2).

6. RESULTS FOR T.N.T.

With the appropriate values of the various constants, the conditions in the detonation wave-front for T.N.T. are determined by the equations

$$p^3 = (171{\cdot}1-43{\cdot}48T/1000)/d, \quad (27')$$

$$v_0' = 5{\cdot}75b+86{\cdot}27/p-6{\cdot}254dp^2, \quad (26')$$

$$v' = v_0'+5{\cdot}75cp+11{\cdot}74dp^2-40{\cdot}47/p. \quad (28')$$

Further, the molecular volume of carbon is 5·4 cm.^{3} and, since at the temperature of the detonation wave-front there are approximately 3·9 moles of carbon, then

$$v_0-v_0' = v-v' = 21. \quad (29')$$

These equations determine the pressure in the detonation wave-front and the detonation velocity as functions of the loading density, once values have been assigned to the virial coefficients. As has been explained above, the values of the virial coefficients have been chosen to give agreement with the measured values of the detonation velocity for loading densities less than 1·5 g.cm.$^{-3}$, corresponding to the lower detonation pressures. With the simple virial expansion which has been assumed for the equation of state it is not possible to get exact correspondence between the experimental and calculated values throughout the whole range of pressure or of detonation velocity. To determine the adiabatic expansion, in which the pressure falls off greatly from its value in the detonation wave-front, it is preferable to have better agreement for lower pressures. This implies that in using the measured values of the detonation velocity to determine the virial coefficients, the equations of state will provide a better approximation if they are chosen to give agreement for low loading densities. The values which are obtained in this way have already been given in equation (7). The detonation pressure and the detonation velocity are shown as functions of loading density in figure 1. The circles represent the experimental determination of the detonation velocity by Friederich (1933, 1936); the departure of the calculated detonation velocity from these values for high loading densities illustrates the remarks made earlier in this paragraph about the determination of the virial coefficients. The values of the pressure show, by reference to table 1 above, that p^*/p is very large in the wave-front and consequently that $N'T/p^*$ is negligibly small, thus justifying the assumptions made in § 5·2 above.

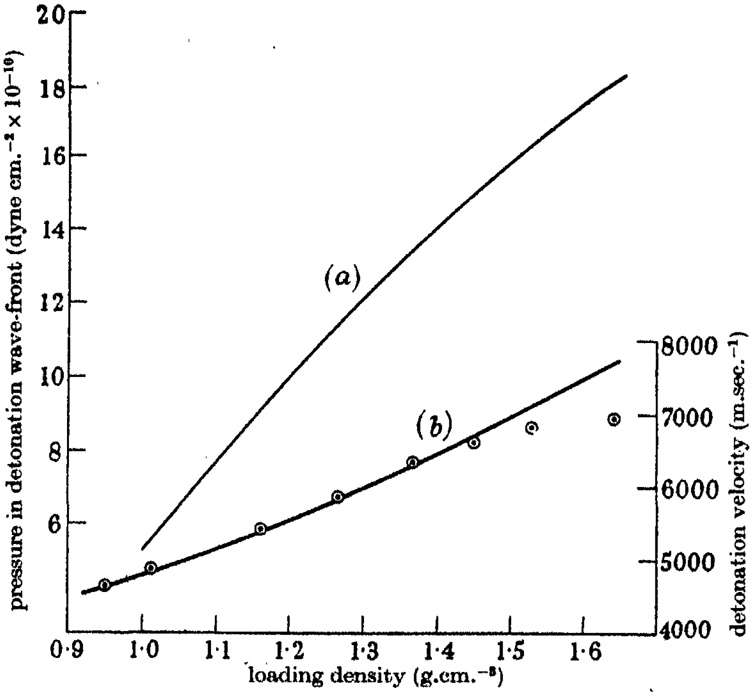


FIGURE 1. (a) Detonation pressure in the wave-front, and (b) detonation velocity as functions of loading density.

PART B. THE ADIABATIC EXPANSION OF THE PRODUCTS OF DETONATION OF T.N.T. AT A HIGH LOADING DENSITY

7. INITIAL CONDITIONS

The initial conditions are determined by the calculations of part A. The equation of state is

$$pv'/N' = \boldsymbol{R}T + 25{\cdot}4p - 0{\cdot}104p^2 + 2{\cdot}33\times10^{-4}p^3, \tag{30}$$

where the pressure is expressed in 1000 kg.cm.$^{-2}$ as unit. For a loading density of 1·5 g.cm.$^{-3}$ the pressure in the detonation wave-front is $15{\cdot}9\times10^{10}$ dyne cm.$^{-2}$, the volume per mole of T.N.T. is 117 cm.^{3} and the temperature is 3400° K. These are the initial conditions for the calculation of the adiabatic relation at this loading density.

8. THE EQUATION FOR ADIABATIC EXPANSION IN A FORM SUITABLE FOR APPLICATION TO SOLID EXPLOSIVES

8·1. The high initial compression of the gases implies both an elaborate equation of state and that the internal energy will depend on the pressure as well as on the temperature. Further, the composition of the products of detonation changes during that part of the adiabatic expansion in which chemical equilibrium is maintained. These two factors make the calculation of the conditions during the adiabatic expansion somewhat complicated. The differential equation for the adiabatic expansion is

$$d(E-H)+p\,dv = 0, \tag{31}$$

where E, which is the internal energy, and H, which is the energy liberated by the chemical reactions, are both complicated functions of pressure and temperature. In order to share these complications between the two terms of equation (31), it has been found advantageous to use T as the independent variable and

$$y = \log_e p^* = \log_e p + (bp + \tfrac{1}{2}cp^2 + \tfrac{1}{3}dp^3)/\boldsymbol{R}T \tag{32}$$

as the dependent variable, for $E-H$ can be expressed more easily in terms of y and T than it can be in terms of p and v.

Equations (8) and (24) can be combined to give

$$E - H = \sum_A N_A E_A + N'E_1 - \sum_{AB} N_{AB}\epsilon_{AB} + \epsilon_{\text{T.N.T.}}, \tag{33}$$

where the first summation is taken over each molecular species and the second summation over all compound molecules, and E_1 is given by equation (9).

8·2. It is convenient to define three functions of p and T by the equations

$$\Phi = \boldsymbol{R}T/(\boldsymbol{R}T + bp + cp^2 + dp^3), \tag{34·1}$$

$$\Psi = (\boldsymbol{R}T + \tfrac{1}{2}cp^2 + \tfrac{2}{3}dp^3)/(\boldsymbol{R}T + bp + cp^2 + dp^3), \tag{34·2}$$

$$X = 1 + y - \log_e p, \tag{34·3}$$

each of which reduces to unity for the perfect gas condition. From these definitions it follows that

$$\left(\frac{\partial p}{\partial y}\right)_T = p\Phi,$$

$$\frac{T}{p}\left(\frac{\partial p}{\partial T}\right)_y = 1-\Psi,$$

$$\left[\frac{\partial(\mathrm{X}T)}{\partial T}\right]_y = \Psi + pv'(1-\Psi)/\boldsymbol{R}TN'.$$

Using these relations it can be shown that

$$\left[\frac{\partial(N'E_1)}{\partial y}\right]_T + p\left(\frac{\partial v'}{\partial y}\right)_T = \mathrm{X}\boldsymbol{R}T\left(\frac{\partial N'}{\partial y}\right)_T - \boldsymbol{R}N'T\Phi \tag{35}$$

and

$$\left[\frac{\partial N'E_1}{\partial T}\right]_y + p\left(\frac{\partial v'}{\partial T}\right)_y = \mathrm{X}\boldsymbol{R}T\left(\frac{\partial N'}{\partial T}\right)_y + \boldsymbol{R}N'\Psi. \tag{36}$$

Equation (31) can be written

$$\left\{\left[\frac{\partial(E-H)}{\partial y}\right]_T + p\left(\frac{\partial v'}{\partial y}\right)_T\right\}dy + \left\{\left[\frac{\partial(E-H)}{\partial T}\right]_y + p\left(\frac{\partial v'}{\partial T}\right)_y\right\}dT = 0,$$

and therefore, using equations (33), (35) and (36), we get

$$-\left(\frac{dy}{dT}\right)_a = \frac{\Sigma\{(E_A-\epsilon_A)(\partial N_A/\partial T)_y + N_A(dE_A/dT)\} + \boldsymbol{R}N'\Psi + \mathrm{X}\boldsymbol{R}T(\partial N'/\partial T)_y}{\Sigma(E_A-\epsilon_A)(\partial N_A/\partial y)_T + \mathrm{X}\boldsymbol{R}T(\partial N'/\partial T)_y - \boldsymbol{R}TN'\Phi}, \tag{37}$$

in which the summations extend over all molecular types and ϵ_A is to be put equal to zero for any chemical elements.

When y has been obtained as a function of T it is easy to obtain p in terms of T from equation (32), $(dp/dT)_a$ from the equation

$$\frac{T}{p}\left(\frac{dp}{dT}\right)_a = T\Phi\left(\frac{dy}{dT}\right)_a + 1 - \Psi \tag{38}$$

and $(dv'/dp)_a$ from the equation

$$-\frac{p}{v'}\left(\frac{dv'}{dp}\right)_a = \Phi\left[1+\frac{p}{\boldsymbol{R}T}\left(\frac{\partial E_1}{\partial p}\right)_T - \frac{p}{T}\left(\frac{\partial T}{\partial p}\right)_a\right]. \tag{39}$$

Equation (37) can be used for any explosive; it will later be reduced to a form suitable for actual calculation in the special case of T.N.T. The effects of the imperfection of the gas and the changing composition can be seen separately in equation (37). If Φ, Ψ and X are each put equal to unity the effect of varying composition in the perfect gas condition is obtained. If $(\partial N_A/\partial T)_y$ and $(\partial N_A/\partial y)_T$ are each put equal to zero for all A, the adiabatic equation for an imperfect gas of fixed composition is obtained. As, in the later stages of expansion, the composition is fixed, the equation for fixed composition will be determined now.

9. EQUATION FOR ADIABATIC EXPANSION WITH FIXED COMPOSITION

When the composition is fixed, equations (37) and (38) reduce to

$$\frac{T}{p}\left(\frac{dp}{dT}\right)_a = 1 + \frac{1}{\boldsymbol{R}}\frac{d}{dT}\left\{\frac{\sum_A N_A E_A}{N'}\right\}, \tag{40}$$

and, since the right-hand side of this equation is a function of temperature only, it can be integrated without difficulty. In this case, equation (39), making use of equation (40), reduces to

$$-\frac{v'}{p}\left(\frac{dp}{dv'}\right)_a = \Gamma\Big/\Phi\left[1+\frac{\Gamma p}{RT}\left(\frac{\partial E_1}{\partial p}\right)_T\right], \tag{41}$$

where $$\Gamma = \left[R+\frac{d}{dT}\left\{\frac{\sum\limits_A N_A E_A}{N'}\right\}\right]\Big/\frac{d}{dT}\left\{\frac{\sum\limits_A N_A E_A}{N'}\right\}.$$

In the limit of low pressures Γ reduces to the ratio of the specific heats at constant pressure and constant volume, and ultimately becomes the exponent of v in the ordinary adiabatic relation. Equation (41) shows that at higher pressures the $\log p$-$\log v$ curve becomes steeper because, although $(\partial E_1/\partial p)_T$ is positive, Φ diminishes rapidly with increasing pressure; this is a consequence of the repulsive forces between the molecules.

Equation (41) can be written in a slightly different form which is of special interest in the case of solid explosives. We have

$$-\left(\frac{dv'}{dp}\right)_a = \frac{N'RT}{p^2}\left[\frac{1}{\Gamma}+\frac{p}{RT}\left(\frac{\partial E_1}{\partial p}\right)_T\right]. \tag{42}$$

The initial velocity of the shock wave in a gas is given by the integral of $(-dv/dp)^{\frac{1}{2}}$ with respect to p from the initial pressure in the detonation wave-front to the pressure in the layer of the shock wave next to the explosive. The effect of the repulsive forces, which is represented by the second member of the right-hand side of equation (42), is to increase the calculated shock wave velocity, and to make p decrease more rapidly with increasing v for low values of v.

10. CALCULATION FOR T.N.T.

10·1. Equations (19·2), (19·3) and (20·1) and the equation obtained by eliminating p^* from equations (20·2) and (20·3) give four equations connecting the relative amounts of water, hydrogen, carbon monoxide, carbon dioxide and methane. Thus, if one of these concentrations be taken as a parameter, say N_{H_2O} which will be written as x, the other concentrations can be expressed in terms of x and T. We then have

$$N_{H_2} = (\theta_1 x^2 + 8{\cdot}5x - 15)/(\theta_3 x - 6) \equiv f(x, T), \tag{43·1}$$

$$N_{CO} = (\theta_1 x^2 + 8{\cdot}5x - 15)/(\theta_2 x - 2{\cdot}5) \equiv g(x, T), \tag{43·2}$$

where $$\theta_1 = \frac{4K_3}{K_1^2 K_2 K_4} - 1, \tag{44·1}$$

$$\theta_2 = 1 - \frac{2}{K_1}, \tag{44·2}$$

$$\theta_3 = 1 - \frac{2K_3}{K_1 K_2 K_4} = \frac{\theta_1+\theta_2}{\theta_2 - 1}, \tag{44·3}$$

and are functions of T only.

The other concentrations then follow from equations (19) and (20) and are given by

$$N_{\mathrm{CO}_2} = 3 - 0{\cdot}5x - 0{\cdot}5g, \tag{45·1}$$

$$N_{\mathrm{CH}_4} = 1{\cdot}25 - 0{\cdot}5x - 0{\cdot}5f, \tag{45·2}$$

$$N_c = 2{\cdot}75 + x + 0{\cdot}5f - 0{\cdot}5g, \tag{45·3}$$

$$N' = 5{\cdot}75 + 0{\cdot}5f + 0{\cdot}5g. \tag{45·4}$$

Five of the six equations (19) and (20) have been used; the remaining one gives a relation

$$y = \log_e p^* = \log_e N_{\mathrm{CO}_2} + \log N' - 2\log_e N_{\mathrm{CO}} + \log_e (K_3/K_4),$$

which can be written as

$$0{\cdot}4343y = \log_{10}(3 - 0{\cdot}5x - 0{\cdot}5g) + \log_{10}(5{\cdot}75 + 0{\cdot}5f + 0{\cdot}5g) - 2\log_{10} g + \log_{10}(K_3/K_4). \tag{46}$$

10·2. At each stage of the integration, that is, for a given pair of values of y and T, it is necessary to determine x from equation (46) using equations (43). Having determined x, f and g, the partial derivatives of f, g and y with respect to x and T, can be calculated. From them, the partial derivatives required in the integration of equation (37) can then be obtained readily.

10·3. If equations (43), (44) and (45) are introduced into equation (37), and the energy terms are grouped together suitably (equations (49) below) the adiabatic equation can be expressed finally as

$$-\left(\frac{dy}{dT}\right)_u = \frac{F(y,T)}{G(y,T)}, \tag{47}$$

where

$$F(y,T) = 5{\cdot}75R\Psi + \frac{d\mathscr{E}_1}{dT} + x\frac{d\mathscr{E}_2}{dT} + f\left\{\frac{d\mathscr{E}_3}{dT} + \tfrac{1}{2}R\Psi\right\} + g\left\{\frac{d\mathscr{E}_4}{dT} + \tfrac{1}{2}R\Psi\right\} + \mathscr{E}_2\left(\frac{\partial x}{\partial T}\right)_y + (\mathscr{E}_3 + \tfrac{1}{2}\mathrm{X}RT)\left(\frac{\partial f}{\partial T}\right)_y + (\mathscr{E}_4 + \tfrac{1}{2}\mathrm{X}RT)\left(\frac{\partial g}{\partial y}\right)_T, \tag{48·1}$$

$$G(y,T) = -(5{\cdot}75 + 0{\cdot}5f + 0{\cdot}5g)\,RT\Phi + \mathscr{E}_2\left(\frac{\partial x}{\partial y}\right)_T + (\mathscr{E}_3 + \tfrac{1}{2}\mathrm{X}RT)\left(\frac{\partial f}{\partial y}\right)_T + (\mathscr{E}_4 + \tfrac{1}{2}\mathrm{X}RT)\left(\frac{\partial g}{\partial y}\right)_T, \tag{48·2}$$

with

$$\mathscr{E}_1(T) = 1{\cdot}25(E_{\mathrm{CH}_4} - \epsilon_{\mathrm{CH}_4}) + 3(E_{\mathrm{CO}_2} - \epsilon_{\mathrm{CO}_2}) + 2{\cdot}75E_{\mathrm{C}} + 1{\cdot}5E_{N_2}, \tag{49·1}$$

$$\mathscr{E}_2(T) = (E_{\mathrm{H}_2\mathrm{O}} - \epsilon_{\mathrm{H}_2\mathrm{O}}) - 0{\cdot}5(E_{\mathrm{CH}_4} - \epsilon_{\mathrm{CH}_4}) - 0{\cdot}5(E_{\mathrm{CO}_2} - \epsilon_{\mathrm{CO}_2}) + 0{\cdot}5E_{\mathrm{C}}, \tag{49·2}$$

$$\mathscr{E}_3(T) = E_{\mathrm{H}_2} - 0{\cdot}5(E_{\mathrm{CH}_4} - \epsilon_{\mathrm{CH}_4}) + 0{\cdot}5E_{\mathrm{C}}, \tag{49·3}$$

$$\mathscr{E}_4(T) = (E_{\mathrm{CO}} - \epsilon_{\mathrm{CO}}) - 0{\cdot}5(E_{\mathrm{CO}_2} - \epsilon_{\mathrm{CO}_2}) - 0{\cdot}5E_{\mathrm{C}}. \tag{49·4}$$

When the values of the functions $\mathscr{E}(T)$ are calculated from the known internal energies and the heats of reaction of the various gases involved and plotted as functions of T, it is found that the points are very nearly collinear. These quantities were therefore represented by linear functions of T as follows:

$$\mathscr{E}_1(T) = 89{\cdot}3T - 343, \tag{50·1}$$

$$\mathscr{E}_2(T) = -0{\cdot}623T - 0{\cdot}513, \tag{50·2}$$

$$\mathscr{E}_3(T) = -1{\cdot}23T + 11{\cdot}5, \tag{50·3}$$

$$\mathscr{E}_4(T) = -2{\cdot}51T + 21{\cdot}9, \tag{50·4}$$

where the temperature is written in terms of 1000° K as unit and the $\mathscr{E}(T)$ are given in kcal. Thus starting from the initial conditions equation (47) can be integrated numerically, the values of the equilibrium constants (and functions of them) being obtained as explained in part A.

10·4. It should be noted that with the given initial conditions

$$p = 3{\cdot}8\,\text{kcal.cm.}^{-3}, \quad T = 3{\cdot}4,$$

equations (32) and (43) to (46) give

$$x = 1{\cdot}102, \quad f = 0{\cdot}0035, \quad g = 0{\cdot}0626.$$

These figures verify in detail that, with the adopted equation of state, only negligible amounts of hydrogen and carbon monoxide are present in the detonation wave-front when the loading density is not less than 1·5 g.cm.$^{-3}$ (cf. the statement following table 2). This is no longer the case for lower loading densities, and the calculation of the adiabatic is then more complicated and has to be carried out as indicated in part C below.

10·5. The integration is continued to the neighbourhood in which there is a minimum amount of free carbon present, which occurs at about 1600° K. From this stage it has been assumed that the composition remains fixed. Using equations (43) to (46) and (49), equation (40) becomes

$$\frac{T}{p}\left(\frac{dp}{dT}\right)_a = 1 + \frac{1}{RN'}\left\{\frac{d\mathscr{E}_1}{dT} + x\frac{d\mathscr{E}_2}{dT} + f\frac{d\mathscr{E}_3}{dT} + g\frac{d\mathscr{E}_4}{dT}\right\}, \tag{51}$$

where $$x = 1{\cdot}049, \quad f = 0{\cdot}3660, \quad g = 1{\cdot}975,$$

their values at 1600° K, calculated in the integration of equation (47). Equation (51) then determines the adiabatic in the region in which the composition remains fixed.

10·6. The concentrations of the products of detonation throughout the expansion are shown in figure 2 and the pressure, volume and temperature relations are shown in table 3. The minimum amount of free carbon present and maximum in the total amount of gas present is evident between 310 cm.^{3} (corresponding to 1800° K) and 410 cm.^{3} (corresponding to 1600° K). In the fourth column of table 3 the values of $-d\log p/d\log v$ are shown. This gives the effective exponent of v in the adiabatic relation during the adiabatic expansion. The high value of this differential coefficient has important consequences which can be seen by considering the variation of $\int p\,dv$ from the initial volume (v_0) up to a given volume (v) with $(v_0/v)^{\frac{1}{4}}$ (which is shown by the broken curve in figure 3). This integral measures the external work done during the expansion, and is equal to the difference between the values of $E-H$ at the

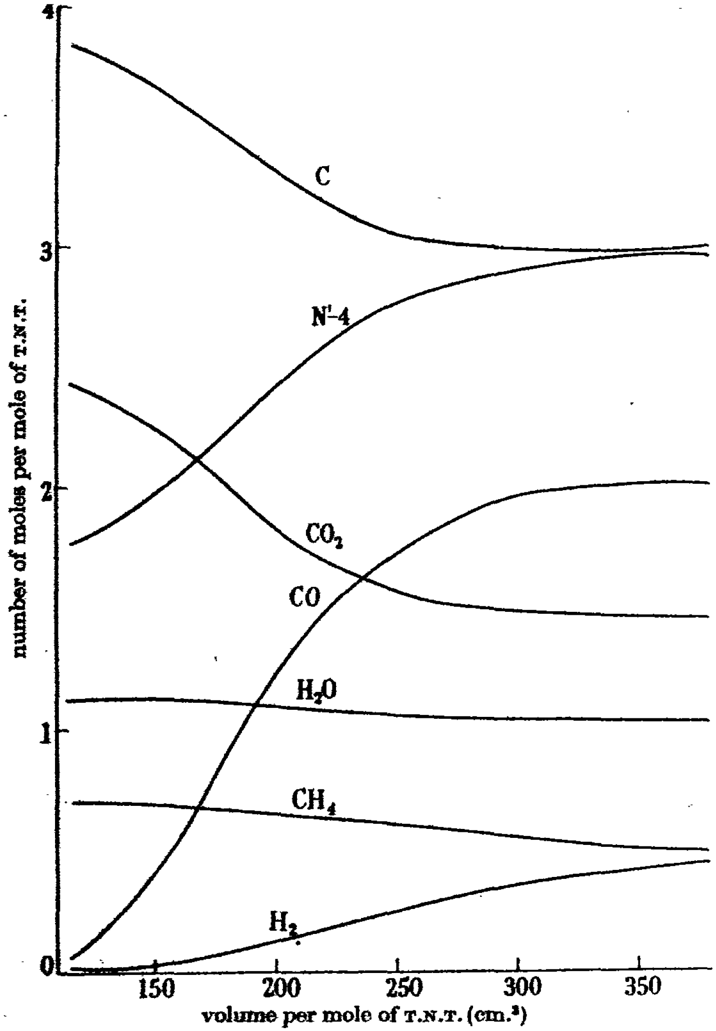


Figure 2. Chemical composition of the products of detonation for a loading density of 1·5 g.cm.$^{-3}$.

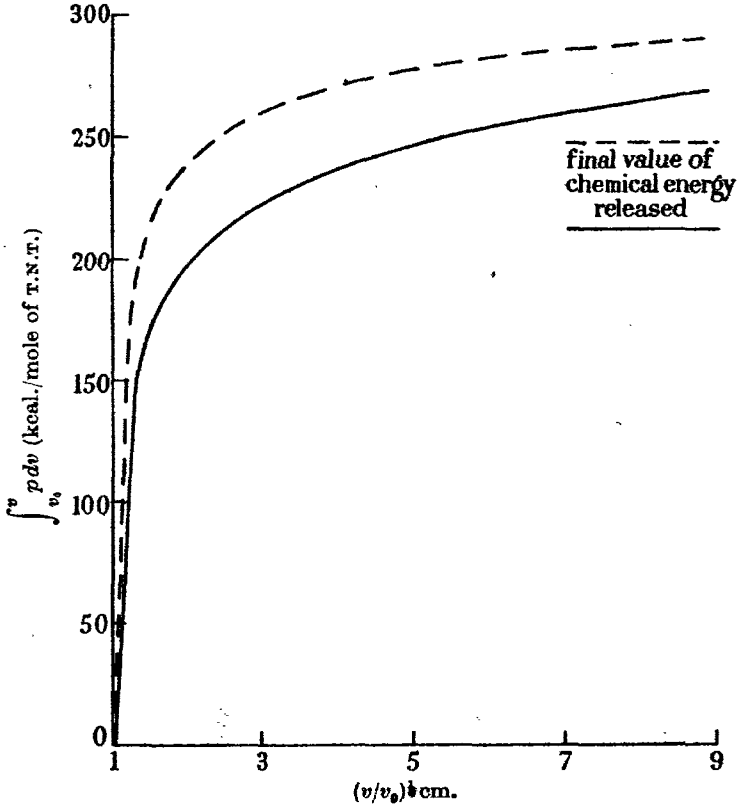


Figure 3. Work done during adiabatic expansion from the initial volume (v_0 cm.^{3}) to a volume v cm.^{3}. The full curve is for a loading density of 1·0 g.cm.$^{-3}$; the broken curve is for a loading density of 1·5 g.cm.$^{-3}$.

initial volume and the volume in question. From figure 3 it can be seen that a large amount of chemical energy is released as work in the very early stages of the expansion. The work done in an adiabatic expansion from a volume v_0 to a volume v, divided by the initial energy of the gas, is

$$1-(v_0/v)^{\gamma-1},$$

where γ is written for $-(d\log p/d\log v)$, and this quantity increases as γ increases. Thus the property of yielding a large part of the chemical energy as work in the early stages of the expansion is associated with the high effective value of γ initially, which corresponds to a sharp fall in pressure for increasing volume for small values of v. As has been noted already this is due to the dominant effect of the repulsive forces between the molecules in the early stages of the expansion.

TABLE 3. TEMPERATURE, PRESSURE AND VOLUME DURING ADIABATIC EXPANSION OF THE DETONATION WAVE-FRONT IN T.N.T. AT A LOADING DENSITY OF 1·5 G./CM.^{3}

temperature (T° K)	pressure (p dyne/cm.^{2})	volume (v cm.^{3}/g.)	γ
3400	15·88 $\times 10^{10}$	0·5136	3·36
3200	10·77	0·5911	2·63
3000	7·809	0·6676	2·76
2800	5·833	0·7421	2·93
2600	4·371	0·8189	3·10
2400	3·238	0·9022	3·27
2200	2·299	1·002	3·30
2000	1·517	1·136	3·05
1800	8·898 $\times 10^9$	1·355	2·76
1700	6·245	1·540	2·39
1600	4·318	1·797	2·09
1500	2·931	2·162	1·86
1400	1·947	2·693	1·68
1300	1·264	3·484	1·54
1200	7·980 $\times 10^8$	4·697	1·44
1100	4·882	6·607	1·37
1000	2·880	9·718	1·32
900	1·626	14·97	1·29
800	8·707 $\times 10^7$	24·29	1·27
700	4·360	41·79	1·27
600	1·999	77·35	1·27
500	8·159 $\times 10^6$	157·0	1·27
400	2·818	362·8	1·27

10·7. For the initial conditions, the chemical energy is 288·6 kcal./mole T.N.T. This is the energy given up in the detonation wave-front, but part of this chemical energy is reabsorbed during the adiabatic expansion as the amount of carbon monoxide increases. When the composition becomes fixed the chemical energy is 247·9 kcal./mole T.N.T. It can be seen from figure 3 that the total work done during the expansion of an element of gas in the detonation wave-front is greater than the chemical energy released. This is because the internal energy of the wave-front is augmented by the work done on it by the detonation wave itself. The maximum kinetic energy in all forms, which can be obtained from the detonation of the explosive is naturally not greater than the final value of the chemical energy released.

11. COMPARISON WITH EXPERIMENT

11·1. Consider an idealized experiment in which the explosive is placed in a large evacuated space, but where the explosive itself is surrounded by a heavy casing. When the explosion takes place under these conditions nearly all the chemical energy is absorbed by the casing, in the early stages of the expansion before the gases have filled the bomb. During this process each element of the products of the explosion is expanding adiabatically except when it passes through the wave-front of a shock wave, and the resulting gases should have very nearly the composition which has been calculated for the adiabatic expansion.

11·2. These ideal conditions are very closely satisfied in the experiments of Robertson & Garner (1923) and in those carried out at the Chemisch-Technische Reichsanstalt (Haid & Schmidt 1929, 1931; Marshall 1932). In the latter experiments, a cartridge of T.N.T. weighing 10 g. at a loading density of 1·52 was fitted into a massive lead block, the end of the hole being filled with sand and the whole placed in an evacuated steel bomb. It was fired with a detonator containing some tetryl. The results of the analysis, compared with the calculations of this paper, are given in table 4 in which the figures give moles of gas per mole of T.N.T. The agreement is fairly good, particularly since in the calculations nitrogen compounds were excluded. In § 15 of part C below, where this question is discussed, evidence will be adduced to show that the nitrogen compounds are probably formed by reactions between the products of detonation while they are cooling to room temperature *after* the expansion has occurred. Thus the occurrence of nitrogen compounds in the experimental analysis indicates that it does not give the products of the adiabatic expansion precisely, and that the values given by it show the effects of chemical reaction after the adiabatic expansion is completed. This probably accounts for differences between the observed and calculated values, particularly in the case of methane.

TABLE 4. DETONATION OF T.N.T. AT A LOADING DENSITY OF 1·5 G.CM.$^{-3}$

	H_2O	CO	CO_2	H_2	CH_4	N_2	HCN	NH_3	C	N'
exp.	1·60	1·99	1·20	0·384	0·007	1·23	0·325	0·211	3·41	6·95
calc.	1·05	1·98	1·49	0·366	0·543	1·50	—	—	2·99	6·92

11·3. The same authors give results of another experiment in which a lead azide detonator is used to fire the charge. The discussion of these results, as also those of Robertson & Garner who carried out experiments at a loading density of 1·3 g.cm.$^{-3}$, can best be deferred until the adiabatic expansion at a low loading density has been considered in part C below.

PART C. THE ADIABATIC EXPANSION OF THE PRODUCTS OF DETONATION OF T.N.T. AT A LOW LOADING DENSITY

12. In the previous part dealing with a high loading density it could be assumed that only negligible amounts of hydrogen and carbon monoxide are present in the detonation wave-front, and this simplified the solution of the equations relating

pressure, volume and temperature. This assumption is no longer valid for a low loading density and this requires that the equations be used differently, as is explained below.

13. THE HUGONIOT CURVE FOR A LOW LOADING DENSITY

13·1. It is assumed that the equation of state (6) holds throughout the expansion with the values of the virial coefficients which have been determined already (equation (30)). The amounts of various substances formed, in terms of the amounts of water, carbon monoxide and hydrogen, are given by equations (43), (44) and (45) and a further equilibrium condition is given by equation (46). For given values of pressure and temperature, equation (30) determines v'/N' and equation (32) determines y. Equations (43) determine f and g for an assumed value of x; these values are then substituted in the right-hand side of equation (46), and the process is repeated until a value of x is chosen which satisfies equation (46). Equations (45) then determine the chemical composition and the total number of moles of gas. This latter value, with the value of v'/N' obtained from equation (30), determines v'. The total volume is then given by

$$v = v' + N_c v_c,$$

where N_c is the number of moles of carbon and v_c, which is equal to 5·4 cm.^{3}, is the volume of 1 mole of carbon.

13·2. The difference between the internal energy, E, and the energy, H, liberated by the chemical reaction in the detonation wave-front is given by

$$E - H = \mathscr{E}_1(T) + x\mathscr{E}_2(T) + f\mathscr{E}_3(T) + g\mathscr{E}_4(T) + \epsilon_{\text{T.N.T.}} + N'E_1, \tag{52}$$

where the $\mathscr{E}(T)$ are given by equations (50) and E_1 is given by equation (9). For a given temperature, the values of $E - H$ and of v can be determined in the way indicated in the preceding paragraph. From the Hugoniot equation,

$$E - H = \tfrac{1}{2}p(v_0 - v), \tag{4}$$

the value of v_0, the initial volume of the explosive, can be determined. For a given temperature, the value of p is varied until this calculation gives a value of v_0 corresponding to the loading density for which it is desired to determine the chemical composition of the products of detonation and the detonation velocity. This determines one point on the Hugoniot curve, and it can be determined completely by carrying through this calculation for a number of different temperatures. By the usual construction of drawing the tangent to this curve from the point on the line $p = 0$ at which $v = v_0$ the detonation velocity can be determined.

13·3. The points which are obtained in this way, assuming that chemical equilibrium is maintained in the wave-front, are plotted as the full curve in figure 4, in which the abscissa is the volume in cm.^{3} and the ordinate is the pressure in 10^{10} dyne cm.$^{-2}$. If α is the inclination of the tangent from the point (227, 0) to this curve, the detonation velocity U m.sec.$^{-1}$ is given by

$$U^2 = 2{\cdot}27 \tan\alpha. \tag{53}$$

This gives a detonation velocity of 5266 m./sec. which is about 5 % greater than the measured detonation velocities at this loading density. The broken curve in figure 4 shows the curve which is obtained when it is assumed that the total number of moles of gas is fixed, and the tangent to this curve (also shown broken) determines the detonation velocity on that assumption. It gives a value about 12 % less than the observed detonation velocity.

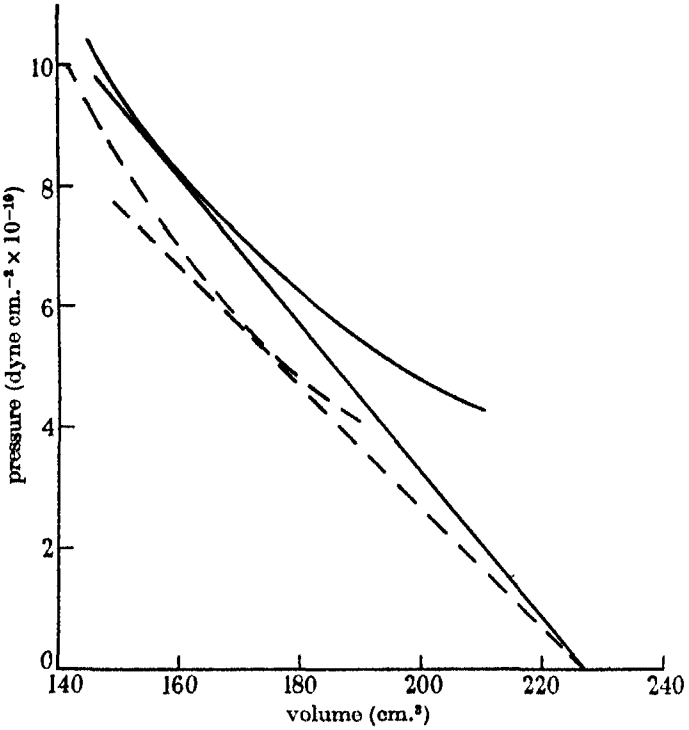


FIGURE 4. Hugoniot curve for a loading density of 1·0 g.cm.$^{-3}$. The full curve is obtained by assuming that chemical equilibrium is maintained in the wave-front, and the broken curve is obtained when the total number of moles of gas is assumed to be constant.

14. THE ADIABATIC EXPANSION OF T.N.T. AT A LOW LOADING DENSITY

14·1. The Hugoniot curve determines the initial conditions for the adiabatic expansion. The starting-point for the integration is given by the values

$$T = 3800^\circ\,\mathrm{K}, \quad p = 8{\cdot}22 \times 10^{10}\,\mathrm{dyne\,cm.}^{-2}, \quad v = 159{\cdot}6\,\mathrm{cm.}^3.$$

$$N_{H_2O} = 0{\cdot}937, \quad N_{H_2} = 0{\cdot}122, \quad N_{CO} = 1{\cdot}286, \quad N_{CO_2} = 1{\cdot}889,$$

$$N_{CH_4} = 0{\cdot}721, \quad N_C = 3{\cdot}104, \quad N' = 6{\cdot}453.$$

From this point, equations (47) and (48) can be integrated numerically. The integration was carried out by determining $(dy/dT)_a$ at the mid-point of each integration interval and using this value of the gradient to determine the value of y at the end of the interval. The pressure and the chemical composition corresponding to these values of y and of the temperature were then determined in the way which has already been indicated (§ 13·1).

14·2. The variation of the chemical composition during the adiabatic expansion is shown in figure 5 and the pressure-volume-temperature relations are given in table 5. From these it can be seen that the amounts of free carbon and of carbon dioxide are a minimum at about 2000° K while the amount of carbon monoxide and the total number of moles of gas are a maximum at about 1950 and 1850° K respectively. Furthermore, for the adiabatic expansion, the composition changes very slowly in relation to volume or pressure in this neighbourhood. It has therefore been assumed (as in part B) that the equilibrium freezes out in this neighbourhood and that thereafter the chemical composition remains fixed. The composition of the gases at this stage (2000° K) is

$$\left.\begin{aligned} N_{H_2O} = 0{\cdot}7372, \quad N_{H_2} = 0{\cdot}6991, \quad N_{CO} = 3{\cdot}5948, \quad N_{CO_2} = 0{\cdot}8340, \\ N_{CH_4} = 0{\cdot}5319, \quad N_C = 2{\cdot}0393, \quad N' = 7{\cdot}897. \end{aligned}\right\} \qquad (54)$$

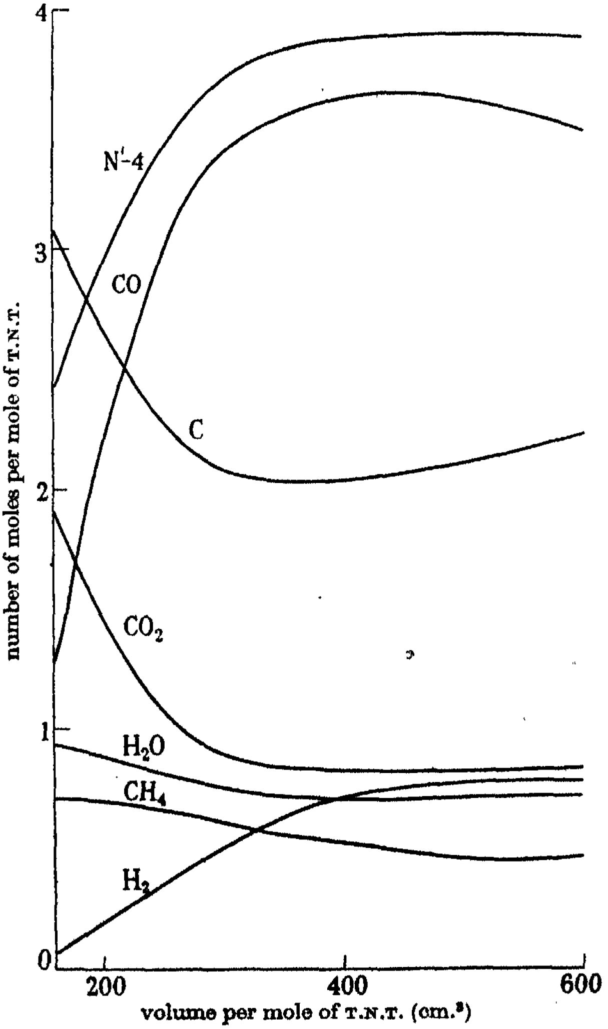


FIGURE 5. Chemical composition of the products of detonation for a loading density of 1·0 g.cm.$^{-3}$.

TABLE 5. TEMPERATURE, PRESSURE AND VOLUME DURING ADIABATIC EXPANSION OF THE DETONATION WAVE-FRONT FOR T.N.T. AT A LOADING DENSITY OF 1·0 G./C.C.

temperature (T° K)	pressure (p dyne/cm.^{2})	volume (v cm.^{3}/g.)	γ
3800	$8{\cdot}224 \times 10^{10}$	0·7035	2·43
3600	6·905	0·7563	2·54
3400	5·759	0·8123	2·69
3200	4·754	0·8727	2·77
3000	3·851	0·9414	2·91
2800	3·045	1·021	2·98
2600	2·334	1·116	2·97
2400	1·726	1·235	2·88
2200	1·191	1·406	2·62
2000	$7{\cdot}420 \times 10^{9}$	1·684	2·22
1800	3·929	2·242	1·90
1700	2·783	2·688	1·73
1600	1·930	3·322	1·59
1500	1·307	4·245	1·48
1400	$8{\cdot}620 \times 10^{8}$	5·624	1·41
1300	5·511	7·752	1·32
1200	3·400	11·13	1·28
1100	2·011	16·76	1·25
1000	1·131	26·50	1·23
900	$5{\cdot}948 \times 10^{7}$	44·69	1·22
800	2·942	79·57	1·21
700	1·314	155·0	1·20
600	$5{\cdot}184 \times 10^{6}$	335·7	1·20
500	1·725	839·3	1·20
400	$4{\cdot}486 \times 10^{5}$	2580	1·20

For a fixed composition the adiabatic is given by equation (41). Substituting the values (54) for the various parameters, this finally reduces to

$$\left(\frac{dp}{dT}\right)_a = 6{\cdot}035\,\frac{p}{T},$$

whence

$$\log_{10} p = \log_{10} T - 2{\cdot}57, \tag{55}$$

in which the constant of integration is determined by the temperature and pressure at 2000° K. Further, for a fixed composition, the volume is given by

$$v = 11{\cdot}01 + 7{\cdot}897(\boldsymbol{R}T + bp + cp^2 + dp^3)/p, \tag{56}$$

in which the constant term is the volume of the solid phase (free carbon). The entries in table 5 for temperatures below 2000° K have been determined from these equations.

14·3. From the calculated values of pressure and volume, the external work done during the expansion has been determined by evaluating $\int p\,dv$ from the initial conditions to any point in the adiabatic expansion. This is plotted as the full curve in figure 3 against the cube root of the ratio of the specific volume at that point to the initial specific volume in the detonation wave-front. By comparing this curve with the broken curve in figure 3 it can be seen that the work done for a loading density of 1·0 g.cm.$^{-3}$ is less than for a loading density of 1·5 g.cm.$^{-3}$. The work done during adiabatic expansion to a given specific volume falls with decreasing loading density on two accounts. First, as the loading density decreases, the specific volume in the detonation wave-front increases so that the ratio v/v_0 falls. Secondly, for a given value of v/v_0 the work done is less for a lower loading density.

14·4. For the initial conditions the amount of chemical energy released in the detonation wave-front is 263·2 kcal./mole of T.N.T.; some of this is reabsorbed as the amount of carbon monoxide increases and when the chemical equilibrium freezes out the amount of chemical energy available is 212·8 kcal./mole of T.N.T. This final value of the chemical energy released is less than the external work done; as already noted in part B, the difference is the amount of work done on the wave-front by the detonation wave itself. By comparing the figures in this paragraph with those given in § 10·7, it will be seen that the chemical energy released per gram of explosive, and, consequently, the kinetic energy which can be obtained from a given mass of explosive is less for a loading density of 1·0 g.cm.$^{-3}$ than it is for a loading density of 1·5 g.cm.$^{-3}$.

14·5. From the calculated values of pressure and volume, the value of $-(d\log p/d\log v)$, more precisely, the value of the ratio $-\Delta\log p/\Delta\log v$, for each integration interval, has been determined, and is given in column 4 of table 5. It is evident that, as in the case of a high loading density, the effective exponent in the adiabatic law is higher in the initial stages of the adiabatic expansion. The behaviour of $-(d\log p/d\log v)$ is the same for low as for high loading densities, and its maximum occurs at very nearly the same volume. The remarks already made in § 10·6 in this connexion apply also in the case of a low loading density.

15. AMOUNTS OF AMMONIA AND OF HYDROCYANIC ACID

15·1. In calculating the adiabatic both for a high loading density and for a low loading density, not all the molecules which could be formed from carbon, nitrogen, hydrogen and oxygen during the detonation have been considered. In particular, the nitrogen has been treated as chemically inactive during the adiabatic expansion. As various amounts of ammonia and of hydrocyanic acid have been reported in experimental investigations (for example, Robertson & Garner 1923), this point needs further investigation. The purpose of this section is to determine the amounts of ammonia and of hydrocyanic acid which are in equilibrium with the other gases under the conditions at which there is a maximum amount of carbon monoxide and total number of moles of gas and a minimum amount of free carbon. The chemical composition under these conditions, which correspond to the freezing of the chemical equilibrium, should correspond to that which is determined in bomb experiments if no reactions occur in the container after the expansion is completed.

15·2. Ammonia and hydrocyanic acid will be formed at the expense of methane, hydrogen, nitrogen and free carbon. The reactions which have to be considered are

$$C\,(\beta\text{ graphite}) + 2H_2 = CH_4, \tag{57·1}$$

$$HCN = \tfrac{1}{2}H_2 + \tfrac{1}{2}N_2 + C\,(\beta\text{ graphite}), \tag{57·2}$$

$$NH_3 = \tfrac{1}{2}N_2 + \tfrac{3}{2}H_2. \tag{57·3}$$

If the equilibrium constants of these reactions at a temperature T° K be denoted by $K_2(T)$, $K_5(T)$ and $K_6(T)$ respectively then, in our usual notation, the chemical composition is determined by the equations

$$N_{\mathrm{NH_3}} + N_{\mathrm{HCN}} + 2N_{\mathrm{N_2}} = 3, \tag{58·1}$$

$$N_{\mathrm{C}} + N_{\mathrm{CH_4}} + N_{\mathrm{CO_2}} + N_{\mathrm{CO}} + N_{\mathrm{HCN}} = 7, \tag{58·2}$$

$$N_{\mathrm{H_2O}} + 2N_{\mathrm{CO_2}} + N_{\mathrm{CO}} = 6 \tag{58·3}$$

and

$$2N_{\mathrm{H_2O}} + 2N_{\mathrm{H_2}} + 4N_{\mathrm{CH_4}} + 3N_{\mathrm{NH_3}} = 5. \tag{58·4}$$

From these equations the values of $N_{\mathrm{N_2}}$, $N_{\mathrm{CO_2}}$, $N_{\mathrm{CH_4}}$, N_{C} and N' can be determined without difficulty in terms of the values of $N_{\mathrm{H_2O}}$, $N_{\mathrm{H_2}}$, N_{CO}, $N_{\mathrm{NH_3}}$ and N_{HCN}. If, following the general equation (16), the equations which express the chemical equilibrium of the reactions (57) be written down and p^*/N' be eliminated from them, simple algebra leads to the equations

$$N^2_{\mathrm{HCN}} = K_5^2 N_{\mathrm{H_2}} (1·5 - 0·5N_{\mathrm{NH_3}} - 0·5N_{\mathrm{HCN}}) \tag{59·1}$$

and

$$K_5 N_{\mathrm{H_2}} N_{\mathrm{NH_3}} = K_2 K_6 N_{\mathrm{HCN}} (1·25 - 0·5N_{\mathrm{H_2O}} - 0·5N_{\mathrm{H_2}} - 0·75N_{\mathrm{NH_3}}). \tag{59·2}$$

15·3. Provided that N_{HCN} and $N_{\mathrm{NH_3}}$ are small, a first approximation to their values can be obtained by putting them equal to zero in the brackets on the right-hand sides of the equations (59). This gives the approximate relations

$$N_{\mathrm{HCN}} = K_5(1·5N_{\mathrm{H_2}})^{\frac{1}{2}}, \tag{60·1}$$

$$N_{\mathrm{NH_3}} = K_2 K_6 \left(\frac{1·5}{N_{\mathrm{H_2}}}\right)^{\frac{1}{2}} (1·25 - 0·5N_{\mathrm{H_2O}} - 0·5N_{\mathrm{H_2}}). \tag{60·2}$$

At 2000° K, using the values given in (54) and the tabulated values of the equilibrium constants (Lewis & von Elbe 1938), these give the values

$$N_{\mathrm{HCN}} = 0·0209, \quad N_{\mathrm{NH_3}} = 0·0093.$$

These values are so small that it is clear that no appreciable error is introduced by replacing the exact equations (59) by the approximate equations (60). In fact, the error in N_{HCN} is not more than 0·5 % and in $N_{\mathrm{NH_3}}$ is not more than 3 %. It is therefore concluded that only minute amounts of ammonia and hydrocyanic acid are formed during the initial stages of the adiabatic expansion and this would not affect the calculations in this paper. It should also be noted that, in discussing the results of their experiments, Robertson & Garner (1923, p. 549) stated that the action of the permanent gases on the liberated carbon during the cooling process determined the amounts of ammonia and cyanogen compounds formed, and that due to continued reaction with the metallic fragments the amount of hydrocyanic acid depends on the rate of removal of the gases from the bomb. The calculations of this section lend support to this conclusion. It is therefore concluded that during the adiabatic expansion only negligibly small amounts of ammonia and hydrocyanic acid are formed, and that the amounts of these substances which have been found in experimental investigations are produced by reactions between the detonation products while they are cooling to room temperature after the adiabatic expansion is completed.

15·4. This also suggests that the analysis of the products of the detonation experiments carried out at the Chemisch-Technische Reichsanstalt (quoted in table 4) does not give the composition at the completion of the adiabatic expansion precisely. Allowing for the fact that most of the ammonia and hydrocyanic acid reported in these experiments was probably formed after the completion of the adiabatic expansion, it appears that the chemical composition calculated in part B for a loading density of 1·5 g.cm.$^{-3}$ is probably in even better agreement with experiment than would appear at first sight.

16. COMPARISON WITH EXPERIMENT

16·1. It has been pointed out already that the conditions which obtain in closed bomb experiments correspond very nearly to those on which the calculations for the adiabatic expansion have been made. In the light of § 15 of this paper, one should add the proviso that precautions must be taken in the experimental investigation to ensure that the analysis of the products give the result of the expansion and not of subsequent reactions. The ratio of the number of moles of carbon monoxide to the number of moles of carbon dioxide is very sensitive to differences in loading density and to the difference in detonation pressure which arises from this. The relevant data are collected in the second and third columns of table 6, together with the available experimental data.

TABLE 6. DEPENDENCE OF N_{CO}/N_{CO_2} ON LOADING DENSITY AND DETONATION PRESSURE

(The unit of pressure is dyne cm.$^{-2}$ × 10^{10})

loading density	1·5 (experimental)	1·5 (calculated)	1·0 (calculated)	1·3 (interpolated)	1·3 (experimental)
pressure when chemical equilibrium freezes out	—	0·888	0·742	—	—
detonation pressure	—	15·9	8·23	—	—
N_{CO}/N_{CO_2}	1·66	1·33	4·31	2·52	2·93

The figures in columns 2 and 3 of this table show that a difference of about 20 % in the pressure at the point in the expansion at which the maximum amount of carbon monoxide occurs changes the ratio N_{CO}/N_{CO_2} by a factor of more than three. The large value of N_{CO}/N_{CO_2} at the lower loading density is almost entirely due to the smaller value of the pressure in the detonation wave-front. In fact, if the detonation pressure, for a loading density of 1·5 g.cm.$^{-3}$, were appreciably less than $15·9 \times 10^{10}$ dyne cm.$^{-2}$, the consequent reduction in the value of the activity would lead to a value of N_{CO}/N_{CO_2} far in excess of the experimentally observed value of 1·66. Although this result depends on the method of calculating the activity (with its various approximations) it does provide some independent evidence for the high value of the pressure in the detonation wave-front as calculated in part A. The value of nearly 16×10^{10} dyne cm.$^{-2}$ is appreciably higher than the values obtained with the co-volume method of Schmidt (1935, 1936) and of Langweiler (1938) who give a value of

$9{\cdot}4 \times 10^{10}$ dyne cm.$^{-2}$ for this loading density. The value 2·52 in the fourth column of table 6 has been obtained from those in columns 2 and 3 by linear interpolation (which is possibly somewhat crude), but it is of interest to compare this figure with that (given in column 5 of table 6) which was obtained by Robertson & Garner (1923) for the ratio of N_{CO}/N_{CO_2} in their experiments at this loading density.

16·2. Some work of interest in connexion with these calculations was carried out at the Chemisch-Technische Reichsanstalt (Haid & Schmidt 1929, 1931; Marshall 1932). The experimental arrangement was as indicated in § 11, and two series of experiments were carried out. In one series a plain detonator consisting of 2 g. of lead azide was used to fire the T.N.T., and in the other a composite detonator which consisted of 0·3 g. of lead azide plus 1 g. of tetryl or 2 g. of lead azide was used to fire the T.N.T., and in the other a composite detonator which consisted of 0·3 g. of lead azide plus 1 g. of tetryl or 2 g. of penthrite; this latter provides a much more powerful initiator. The chemical composition of the products of the detonation in the two cases are given in table 7; the figures are the number of moles of the different products per kg. of T.N.T.

TABLE 7. PRODUCTS OF DETONATION OF T.N.T. USING DIFFERENT INITIATORS

detonator	plain	composite
loading density	1·52	1·52
CO_2	1·78	5·30
CO	18·63	8·79
H_2O	4·25	7·05
H_2	5·34	1·69
CH_4	0·10	0·03
HCN	1·1	1·43
C_2N_2	1·16	—
NH_3	0·33	0·93
N_2	4·74	5·42
C	6·6	15

Although the loading density was the same in the two cases there is an enormous variation in the chemical composition of the products of detonation; in particular, with the plain detonator there is 10·5 times as much carbon monoxide as carbon dioxide, while with the composite detonator this ratio is 1·66. The calculations of this and the preceding part provide an explanation of this. The relatively weak lead azide detonator fails to produce complete decomposition in the shock wave-front of the detonation zone. As a result, the pressure moves ahead of the chemical decomposition, as in the case of a laterally expanding linear charge, and therefore expands the cavity in the lead container enclosing the charge. Thus for a considerable fraction of the charge of T.N.T., the effective loading density at which chemical decomposition takes place is very much reduced. Consequently, on the basis of the calculations in this and the preceding part, a large increase in the value of the ratio of the number of moles of carbon monoxide to the number of moles of carbon dioxide is to be expected; and this is, in fact, observed. It appears that this effect could be made the basis of a means for measuring the effectiveness of initiators, since the value of the ratio N_{CO}/N_{CO_2}, which can be determined experimentally, gives a measure of the extent to which the pressure can get ahead of the chemical decomposition in the detonation resulting from the use of any particular initiator, which in turn determines the effectiveness of a lightly cased charge.

The work described in this paper was carried out for the Research and Experiments branch of the Ministry of Home Security to whom our thanks are due for permission to publish. The substance of it was contained in reports numbers R.C. 166 dated January 1941, R.C. 212 dated June 1941 and R.C. 306 dated February 1942. It is a pleasure to thank Sir Geoffrey Taylor for his interest in the work described in this paper.

REFERENCES

Cranz 1926 *Lehrbuch der Ballistik*, **2**, 107.
Crow & Grimshaw 1930 *Phil. Trans.* A, **230**, 39.
Friederich 1933 *Z. ges. Schiess- u. Sprengstoffw.* **28**, 2, 51, 80, 113.
Friederich 1936 *Z. ges. Schiess- u. Sprengstoffw.* **31**, 252, 288.
Garner & Abernethy 1921 *Proc. Roy. Soc.* A, **99**, 213.
Garner & Abernethy 1930 *J. Amer. Chem. Soc.* **52**, 116.
Haid & Schmidt 1929 *Jber. chem.-tech. Reichsanst., Lpz.*, **8**, 128.
Haid & Schmidt 1931 *Z. ges. Schiess- u. Sprengstoffw.* **26**, 253.
Jouguet 1906 *J. Math. pures appl.* (2), **6**, 5.
Langweiler 1938 *Z. techn. Phys.* **19**, 271.
Lewis & von Elbe 1938 *Combustion, Flames, and Explosions in gases.* Cambridge University Press.
Marshall 1932 *Explosives*, 2nd ed., **3**, 141. London: J. and A. Churchill.
Robertson & Garner 1923 *Proc. Roy. Soc.* A, **103**, 539.
Schmidt 1935 *Z. ges. Schiess- u. Sprengstoffw.* **30**, 364.
Schmidt 1936 *Z. ges. Schiess- u. Sprengstoffw.* **31**, 322.

A study of expected loss rates in the counting of particles from pulsed sources

By C. H. Westcott, Ph.D., *Nuffield Research Fellow, Physics Department, Birmingham University*

(*Communicated by M. L. E. Oliphant, F.R.S.—Received* 7 *January* 1948)

The use in nuclear physics of accelerators giving a pulsed output leads to difficulties when electrical methods are used for detecting the particles produced. The counting losses due to the finite resolving time of the counting system are enhanced by reason of the pulsed nature of the source, and may become considerable unless the counting speed is kept quite low. The present paper contains some calculations on the loss rates to be expected. It is desirable for the resolving time of the counting system, i.e. the dead time following each count, to be much shorter than the pulse length of the accelerator, but since this is itself usually only a few microseconds, this condition is not easy to achieve. If the dead-time is greater than the duration of the accelerator pulse, the calculations are relatively easy but the losses may be high. Calculations for intermediate cases show that the losses can be estimated fairly accurately, and results of practical value can be obtained up to fairly high counting speeds when the dead-time is as high as 40 % of the pulse length. The accuracy with which the loss can be calculated will usually be limited by the uncertainty in our knowledge of the shape of the output pulse of the accelerator and of the exact length of the dead-time of the counting arrangement.

The attention of the reader is particularly directed to appendix III, where the results of the calculations of this paper will be found summarized. The experimental physicist who wishes to make use of the results without following the detailed analysis may pass directly from the end of § 1 to this appendix, which will also be of value to others for rapid reference.

Nomenclature and list of symbols

Nomenclature has proved a difficulty in writing this paper. The word 'pulse' has been reserved throughout for the pulse given by the accelerator; the signal in the counting circuits caused by the particles being detected is called a 'count' whether or not it is recorded by the scaler or other instrument used. Thus there are 'suppressed counts' and 'recorded counts'. Readers should distinguish between two probabilities, viz. the probability that, if a count occurs, it shall be recorded, which may be 85 %, or the probability per unit time of there being a recorded count (which may be 1 % per μsec.).

p = recurrence frequency of pulses.
$\mathscr{T}$ = pulse length. $f = p\mathscr{T}$ = fractional pulse duration.
t = time. T = total duration of counting.
τ = resolving time of system, or 'dead-time' following a count.
ν = mean counting rate during pulse.
μ = mean counting rate overall. $\mu = f\nu$.
E = μT = expected total count (including suppressed counts).
E' = expected total of recorded counts.
τ' = effective dead-time from first-order loss rate.
$P(t)$ = probability (as a function of time) that a count will be recorded.
Π = probability of counting within a pulse period.

x $= t/\tau =$ time in units of the dead-time.

X $= \mathcal{T}/\tau =$ duration of the pulse in units of the dead-time.

z $= \nu\tau =$ mean counting rate (during the pulse) in units of the dead-time, or average expectation of counts within one dead-time.

$N(t) = \int_0^t \nu(t)\,dt$, when ν is variable.

$\nu(t)$ $= k\nu_0(t)$. $\nu_0(t)$ is a form factor (mean value over pulse = unity); k is a parameter representing counting speed ($k = \bar{\nu}$).

μ_0, E_0, E'_0, etc., as μ, E, E' but referring to conditions when only the background counts are present.

E'_n refers to conditions when there is no background outside the pulse, but the background rate is added during the pulse.

1. INTRODUCTION

One of the outstanding trends in the present phase of research in nuclear physics is the use of increasingly large and complex machines for the acceleration of charged particles to very high energies. Results of fundamental importance are to be expected from developments in this field, although many of the techniques for dealing with these high-energy particles have yet to be fully worked out. A feature common to all machines giving particles of the highest energies is that their output is not continuous, but occurs in bursts or 'pulses' separated by relatively long intervals during which the machine gives no output. For example, to overcome the limitations imposed by the relativistic increase of mass with velocity, cyclotrons designed to give particles of velocities above about 0·15c generally use a frequency-modulation technique, and therefore give a pulsed output. Betatrons and synchrotrons do the same, it being inherent in their design that the particles only emerge when the magnetic field has reached a predetermined value. Linear accelerators using the micro-wave technique also give a pulsed output of particles, for they are usually powered from pulsed magnetrons of the type developed for radar, and in any case considerations of power dissipation in the accelerator alone preclude continuous working.

The pulsed nature of the source is of no consequence when it is used to manufacture radioactive isotopes, even if these are subsequently to be studied with counters—the amount of radioactive substance produced will clearly depend only on the mean beam current and the exposure time. Equally, when it is convenient to use the tracks generated in a photographic emulsion for studying the reactions, the pulsed nature of the source is irrelevant. For use with a cloud chamber it may even be an advantage, especially if the pulses are quite infrequent, for one can arrange to expand the chamber synchronously with the pulses of the machine. There remains, however, a large field of investigation where quantitative results are of importance and where methods involving electrical counting of the instantaneous products of the reaction are pre-eminently suitable. On account of statistical fluctuations it becomes necessary to observe a large number of events (e.g. ten thousand if a 1 % accuracy is required); to do this with a cloud chamber is out of the question, and with the photographic track technique is usually very tedious. Within this field are to be included observations of scattering laws, both elastic and inelastic, as well as many types of reactions involving the emission of charged particles. High accuracy is always desirable for the determination of cross-sections and for measuring ranges of product particles when several overlapping groups occur, and in determining the fundamental scattering laws it is particularly necessary. It is just in this field of investigation that a characteristic disadvantage of using a pulsed source becomes important; moreover, it is a disadvantage which has hitherto passed relatively unnoticed.

The difficulty which arises in electrical counting when the source is pulsed is due to the finite resolution of any counting system. Thus, following any count, there is a dead-time during which the system would be unable to record a further count, should one occur. Calculations on the effect of this when the source of particles is of constant intensity have been made by Locher (1933), Volz (1935), Skinner (1935) and Ruark & Brammer (1937). These writers were primarily concerned with Geiger-Müller counters, for which the dead-time usually exceeds $100\,\mu$sec. For systems using proportional counters or ionization chambers with linear amplifiers, much shorter dead-times can be achieved. Nevertheless, as will be seen below, with pulsed sources the losses may become large even with linear amplifier systems, and Geiger counters are generally out of the question for this work owing to the magnitude of the losses which would be expected. Thus to obtain accurate results in a reasonable time, we may wish to count at an average rate of, say, 1000 per min. If the duty cycle of the machine is 1/1000 (e.g. $5\,\mu$sec. pulses every 5 msec.), this means that, during the pulse, particles must be arriving at a mean rate of a million per minute. It is this fact which makes conditions so stringent.

In the present paper a careful study is made of the counting losses expected under these conditions. It is seen that one cannot simply apply the equations for continuous counting at the enhanced counting speed; the error in doing so would be greatest for conditions which are very likely to arise in practice, i.e. when the pulse length does not greatly exceed the dead-time of the system. It is also shown that it is desirable to reduce the dead-time to the minimum, the values required for good results lying quite close to the limit practicable with present electronic technique. Indeed, a few of the accelerating machines now planned, using pulsed deflector systems, will give an output pulse only a few hundredths of a microsecond long; for use with these, present electronic technique already requires improvement if anything but slow counting rates are to be used.

2. FUNDAMENTAL ASSUMPTIONS

As was pointed out by Ruark & Brammer (1937), the dead-time mechanism may take one of two forms. The inoperative period may follow only counts which are recorded, or it may follow all counts, including those suppressed by the dead period after a previous count, thus lengthening the total dead period until a time τ after the last count to arrive (figure 1). The latter is the case, for example, in the Geiger counter, where the suppression is internal and due to an ion sheath, but the duration of the dead period usually depends on the counting rate, as does the amplitude of the counter output. In practice, Geiger counters are seldom used with pulsed sources, on account of their long inherent dead-time. When working with pulsed sources of particles, it is preferable to use either proportional counters or ionization chambers with linear amplifiers. The dead-time of such systems will depend on the circuits used; for example, on the discriminator level and on the amplitude and wave form of the counting signal when using a 'straight' discriminator, without a relaxation time of its own. In conditions where a calculation of counting losses is important, it seems preferable to insert electronically a definite dead-time slightly greater than the largest inherent dead-time of the system, so that variations of the latter become unimportant. As usually arranged, the electronic dead-time follows recorded counts only. It is true that to do this a careful electronic technique is needed when dead-times of only about 1 μsec. are required, but the importance of having an accurately known and constant dead-time usually justifies the trouble involved. In the present calculations, therefore, a constant dead-time τ to follow each *recorded* count is assumed.

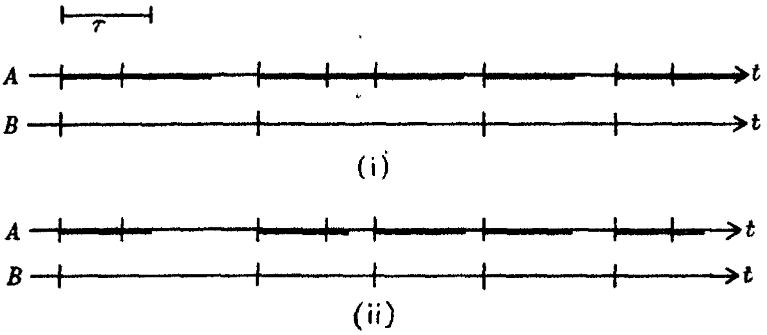


FIGURE 1. Illustrating the two types of dead-time mechanism. The upper line (A) shows all counts and the lower line (B) only the recorded counts. The double horizontal lines indicate the duration of the dead-times (cf. Hole 1946). (i) Dead-time follows all counts. (ii) Dead-time follows recorded counts only.

We also follow Bateman (1910) and Marsden & Barratt (1911) in assuming that the original counts occur at random. Thus if the expected counting rate throughout any period is μ, the probability $p(t)\,dt$ that a count will occur between time t and $t+dt$ from any arbitrary zero of time is just $\mu\,dt$, and the probability that the first count after zero time shall thus occur is $p_1(t)\,dt = \mu e^{-\mu t}\,dt$. Further, the probability that the nth count shall occur between t and $t+dt$ is

$$p_n(t)\,dt = \frac{\mu^n t^{n-1}}{(n-1)!} e^{-\mu t}\,dt, \tag{1}$$

and the probability of finding just n particles in an interval of length t is $(\mu t)^n e^{-\mu t}/n!$. These results still apply if the zero of time is identified with a particular count, so that, for example, the probability of finding no counts within an interval t after any count is just $e^{-\mu t}$. This is, of course, the same as the probability that the interval between any two successive counts is greater than t.

For the case of continuous counting with a uniform expected counting rate μ, therefore, we see that the probability of any count being less than τ after its immediate predecessor is $1-e^{-\mu\tau}$, so if the dead-time followed all counts (whether or not recorded) the expected recorded count would be $E' = Ee^{-\mu\tau}$, where $E = \mu T$ is the total count expected in time T, including suppressed counts. However, when the dead-time follows only recorded counts, we proceed otherwise; noting that the total dead-time is $E'\tau$, and that in the remaining time $T-E'\tau$ the probability of a recorded count is the same as that of any count occurring (viz. μdt in any dt), we see that $E' = \mu(T-E'\tau)$, so that

$$E' = \frac{\mu T}{1+\mu\tau} = \frac{E}{1+\mu\tau}. \tag{2}$$

We may note here that, to a first order in $\mu\tau$, the result is the same whether the dead-time follows all counts or only recorded counts, being $E' = E(1-\mu\tau+O(\mu^2\tau^2))$. The above treatment also neglects statistical fluctuations, in the sense that E' is only an expected count. The actual count will fluctuate somewhat, the standard deviation on E being $\pm\sqrt{E}$; in fact, the standard deviation on E' should be slightly less than its square root, although the difference is small when the fractional loss $1-E'/E$ is small, and this point will be referred to later. It is a measure of the fact that the experimental result gives less accurate information than it would if no counting losses occurred, $\frac{\partial E'}{\partial \mu}$ being less than $\frac{E'}{E}\frac{\partial E}{\partial \mu}$.

It is convenient to give a numerical example here. If $\tau = 10^{-4}$ sec. and $\mu = 100$ per sec. (6000 counts per min.), $\mu\tau = 10^{-2}$ and the loss rate is only 1 %; at 60,000 counts per min. it would be about 10 %. We shall see below that with pulsed sources conditions are much more stringent.

3. Calculations for a pulse source without background count

For simplicity we begin by assuming that the source gives pulses of length $\mathcal{T}$ at a uniform rate of p per unit time. Within the pulse period we assume that its intensity is constant, giving an expected counting rate ν. We further assume that there is no background count, i.e. that the counting rate between pulses is zero. The overall mean true counting rate is, of course, $\mu = \nu f$, where $f = p\mathcal{T}$, and we further assume f is small, so that the interval $(1/p-\mathcal{T})$ between pulses is much greater than $\mathcal{T}$. We also assume $1/p-\mathcal{T} > \tau$, so that no count recorded in one pulse can affect conditions in the next. These last two conditions are in practice almost always satisfied; the first two are not, and we shall deal with them later in this paper.

The approach to these calculations is somewhat different according to the value of τ. We shall consider three cases, viz. $\tau > \mathcal{T}$, $\tau \ll \mathcal{T}$, and the intermediate case.

3·1. *Dead-time greater than the pulse length* ($\tau > \mathcal{T}$)

In this case only one count can be recorded in any one pulse. The probability of there being no count within the pulse is simply $e^{-\nu\mathcal{T}}$; in all other cases there is one recorded count, the probability of this being, of course, $1-e^{-\nu\mathcal{T}}$. In total time T

there are pT pulses, and the statistical distribution is of the binomial or Bernouilli type; in particular, the expected total recorded count is given by

$$E' = pT(1-e^{-\nu\mathscr{T}}) = pT(\nu\mathscr{T} - \tfrac{1}{2}\nu^2\mathscr{T}^2 + \tfrac{1}{6}\nu^3\mathscr{T}^3 - \dots) = \nu fT(1 - \tfrac{1}{2}\nu\mathscr{T} + \tfrac{1}{6}\nu^2\mathscr{T}^2 - \dots). \quad (3)$$

We see that this expression is independent of τ (as long as $1/p - \mathscr{T} > \tau > \mathscr{T}$), and the first-order correction term is a fraction $\frac{1}{2}\nu\mathscr{T} = \mu\mathscr{T}/(2f) = \mu/(2p)$ of the total count. Comparing this with the formula for steady counting at rate μ, we see that the effective dead-time of such a system is given by

$$\tau' = 1/(2p), \quad (4)$$

or half the pulse spacing. This may be very much longer than the actual dead-time, so that we see that such a system may be very poor at high counting rates.

To take a numerical example which will be more or less typical, we may put $p = 100$ pulses per sec., $\mathscr{T} = 5\,\mu$sec.; then as soon as τ exceeds $5\,\mu$sec., τ' becomes 5 msec., and at a mean counting rate of 500 per min. we have already a 4 % counting loss. This is clearly much worse than when the counts are not bunched into pulses; we may compare these figures with the example given in § 2, i.e. 1 % loss at 6000 per min. for $\tau = 10^{-4}$ sec., or 1 % loss at 60,000 per min. for $\tau = 10^{-5}$ sec.

The fact that the standard deviation of the total recorded count decreases when losses become appreciable, already mentioned in § 2, is well illustrated by this case. For the binomial distribution the standard deviation is $\sqrt{\{pt(1-e^{-\nu\mathscr{T}})\,e^{-\nu\mathscr{T}}\}}$, or $e^{-\frac{1}{2}\nu\mathscr{T}}$ less than the square root of the total count. However large the true counting rate, the number of counts recorded in a second can never exceed p, and as this saturation condition is approached the standard deviation of the observed count drops; at the same time the estimate of the true count deduced therefrom becomes increasingly inaccurate. Thus in the somewhat extreme case when the true counting rate is 2 per pulse (say $p = 100$, $\mu = 200$, $\nu\mathscr{T} = 2$), the probability of a pulse not giving rise to a count is only 13·5 %; the expected recorded count in a unit time is $86\frac{1}{2}$, but the statistical error is more nearly $\sqrt{13\frac{1}{2}}$ than $\sqrt{86\frac{1}{2}}$, being in fact $\sqrt{11{\cdot}7}$ from the above formula.

3·2. *Dead-time much shorter than the pulse length*

To avoid these restrictions it is desirable, whenever even medium counting speeds are desired, to make the dead-time shorter than the pulse length. If it can be made much shorter, the statistical conditions are simplified, but this is not always easy to achieve; thus if $\mathscr{T} = 5\,\mu$sec., a dead-time of 10^{-7} sec. or less is required.

This case is most easily studied by comparing a period T of counting of the pulsed source with a period fT of counting at a steady rate ν. The same true count is to be expected, the effect being to telescope the pulse periods together by removing the intervals. In doing this, some additional counts will be suppressed, but these losses are confined to the first τ of each pulse, being due to counts suppressed by other counts in the last τ of the previous pulse. The mean count expected in either the first or the last τ of any pulse is $\nu\tau$, and the counts at the end of a pulse are equally likely to occur at any time within the period τ, so that on the average they will suppress half the counts in the first τ of the next pulse when the pulses are telescoped (figure 2). The loss on telescoping therefore averages $\frac{1}{2}\nu^2\tau^2$ per pulse, while the expected recorded count per pulse after telescoping is $\nu\mathscr{T}/(1+\nu\tau)$, so that the counting rate recorded in the actual case should be

$$pT\nu\mathscr{T}\left(\frac{1}{1+\nu\tau}+\frac{\nu\tau^2}{2\mathscr{T}}\right) \doteqdot \mu T\left[1-\mu\left(1-\frac{\tau}{2\mathscr{T}}\right)\frac{\tau}{f}\right]. \tag{5}$$

This result is only taken to the first order in $\nu\tau$, but since in this case τ is of the order of 10^{-7} sec. or shorter, the limitation is unlikely to be important. We see from this formula that the effective dead-time τ' is given by

$$\tau' = \frac{\tau}{f}\left(1-\frac{1}{2}\frac{\tau}{\mathscr{T}}\right), \tag{6}$$

being for $\tau \ll \mathscr{T}$ approximately τ/f. Thus for $p = 100$, $\mathscr{T} = 5\,\mu$sec. and $\tau = 10^{-7}$ sec., τ' comes to $1{\cdot}98\times10^{-4}$ sec., or 1980 times the actual dead-time. Counting losses at 6000 counts per min. are for this case about 2 %.

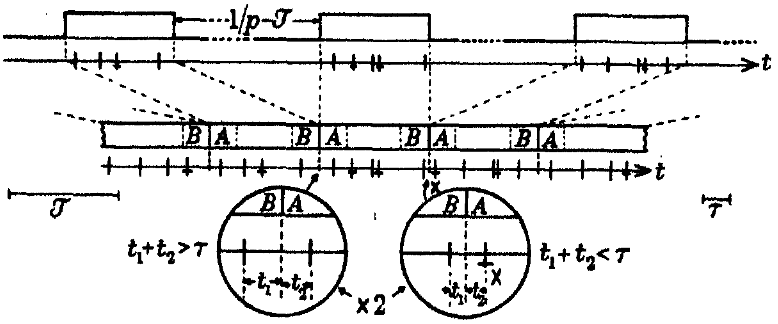


FIGURE 2. Illustrating the 'telescoping' of pulses. Counts are indicated by vertical lines across the t-axis; suppressed counts are crossed below the axis. The count marked X would be suppressed in the telescoping process—see inset enlarged portions of the diagram. Suppression of this kind occurs if $t_1+t_2<\tau$. A count in period B suppresses any count in the first $\tau-t_1$ of the following pulse, but $\bar{t}_1 = \frac{1}{2}\tau$, so that average count in B suppresses counts in a period $\frac{1}{2}\tau$ in the next pulse. (N.B. This method is applied when $\tau \ll \mathscr{T}$ and $\nu\tau$ is small; to simplify the drawing these conditions are not satisfied in the above figure.)

3·3. *More exact results for* $\tau < \mathscr{T}$

When $\mathscr{T}/\tau$, although greater than unity, cannot be considered large, more detailed calculations are necessary. This case is of considerable practical importance in that resolving times much less than $1\,\mu$sec. are likely to be difficult to achieve. We may note at the outset that so long as the counting rate is so small that terms in $(\nu\tau)^2$ can be neglected, the analysis of § 3·2 above holds, nor does it matter whether the dead-time follows all counts or only recorded ones. It is, however, desirable to push the analysis further to see how high a counting rate may be used before the correction terms become too big or too difficult to estimate. We may also notice in passing that for the case of $\tau = \mathscr{T}$ and $\nu\tau \ll 1$, the analyses of §§ 3·1 and 3·2 above both give the same first-order result, viz. $E' = \mu T(1-\frac{1}{2}\nu\tau)$.

It is convenient for the further analysis to change our variables and take the dead-time τ as the new unit of time. We then write $x = t/\tau$ as the time variable, taking it as zero at the beginning of any pulse; writing $X = \mathscr{T}/\tau$, we see that the pulse extends from $x = 0$ to $x = X$. We also write $z = \nu\tau$ as the new variable describing the counting speed during the pulse; z is just the mean number of counts expected to occur within one dead-time. The probability of a count occurring in dt is, of course, νdt $(0 < t < \mathcal{T})$, and since $dt = \tau dx$ we see that this becomes $z dx$ in our new units.

We next require the probability, $P(x)$ say, that a count, having occurred, shall be recorded. We may note that $P(0) = 1$, since the interval between pulses is longer than the dead-time, so that the first count in a pulse is never suppressed. Further, if the pulse is long enough, $P(x)$ settles down for large x to the value $1/(1+z)$ which characterizes continuous counting at rate z. If we know P, we can readily calculate the total count in time T; that is, in fact, just

$$E' = pT\int_0^X P(x)\, z\, dx = E/X\int_0^X P(x)\, dx = E\,\overline{P(x)}, \tag{7}$$

where the bar indicates a mean value over the pulse time.

For the range $0 < x < 1$, P is readily calculated, since the first count always records, and no other count can record within the dead-time. The probability of the first count falling within dx at x being $ze^{-zx}dx$, it follows that

$$P(x) = e^{-zx} \quad (0 < x < 1). \tag{8}$$

For the range $1 < x < 2$, P is the sum of two terms; $P(x)\, z\, dx = ze^{-zx}dx$ as above *plus* the probability of the second recorded count falling in dx at x. For this to occur there will have been one complete dead-time and only one between 0 and x, so that for a unit period there was no probability of a recorded count, while for a period totalling $x-1$ there was a probability $z dx$ in every dx (since every count occurring outside a dead-time records). The probability of the second recorded count falling within dx at x is therefore the same as the probability of the second count falling within dx at $x-1$ if there were no dead-time, which is (cf. equation (1))

$$z^2(x-1)\, e^{-z(x-1)}\, dx.*$$

We have, therefore, that, in the range $1 < x < 2$,

$$P(x) = e^{-zx} + z(x-1)\, e^{-z(x-1)}. \tag{9}$$

Proceeding similarly, we obtain for $2 < x < 3$,

$$P(x) = e^{-zx} + z(x-1)\, e^{-z(x-1)} + \tfrac{1}{2}z^2(x-2)^2\, e^{-z(x-2)}, \tag{10}$$

and clearly we can write down similar expressions for later ranges of x.

Figure 3 is a graph of $P(x)$ for values of x from 0 to 3 and for various values of z. In practice, values of z of much over 0·5 are unlikely to be used, but the curve for

---

* An alternative proof is as follows: The second recorded count can only occur at x if the first occurs between 0 and $x-1$, in dy at y, say. The probability of this is $ze^{-zy}dy$. Then no count can occur until after $y+1$; take this as a new zero of time, and the next count must occur after an interval $x-(y+1)$ and within a range dx if $dy \ll dx$. The probability of this is $ze^{-z(x-y-1)}dx$; the two probabilities are independent and may therefore be multiplied. Integrating with respect to y, which involves proceeding to the limit $dy \to 0$, and therefore automatically satisfies $dy \ll dx$, we obtain the probability of the second recorded count being within dx at x as

$$\int_0^{x-1} ze^{-zy}\, ze^{-z(x-y-1)}\, dx\, dy = z^2 e^{-z(x-1)}\, dx \int_0^{x-1} dy = z^2(x-1)\, e^{-z(x-1)}\, dx.$$

$z = 2{\cdot}5$ is interesting in that it exhibits peaks just after $x = 1$ and $x = 2$; we should expect this, for with such a high value of z the first count is likely to occur very soon after the start of the pulse, and therefore the dead-time ends just after $x = 1$ and a second recorded count is expected very soon after this; the second dead-time ends soon after $x = 2$ and a third count follows almost immediately. This illustrates the fact that for very high values of z this system tends to give regular counts spaced just τ apart with practically no statistical fluctuations. Such a result is, of course, useless in practice when it is an estimate of z which is required.

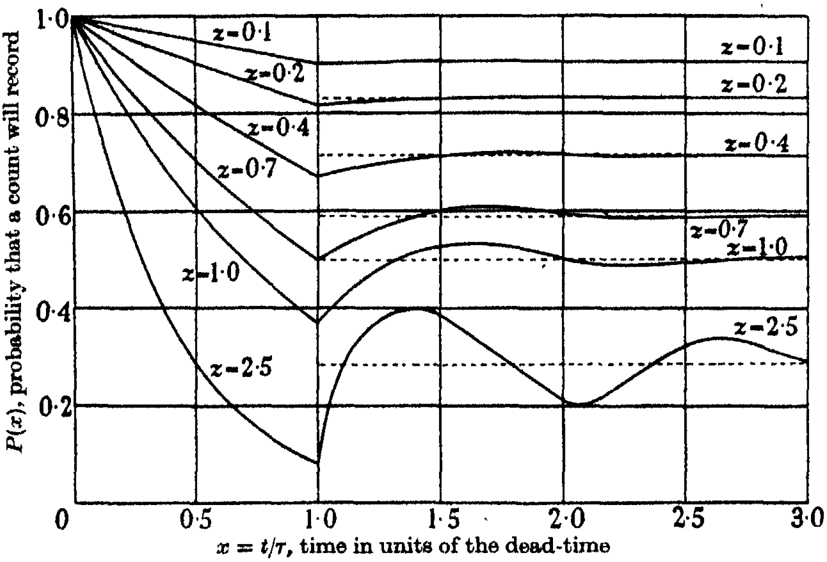


FIGURE 3. Probability that a count records as a function of time since the pulse began. The parameter z represents mean counting rate. The dotted lines are given by $P = 1/(1+z)$, to which the actual curves are asymptotic for large x. In approximation B these dotted curves are followed for all $x > 1$; for approximation C for all $x > 2$; and for approximation D the full curves are used for $x < 3$ and the value $1/(1+z)$ thereafter.

It is seen from figure 3 that for values of z up to $0{\cdot}7$ at least, P converges very rapidly towards its final value of $1/(1+z)$, and for almost all practical cases this expression may be used for $P(x)$ whenever $x > 2$; even to do so whenever $x > 1$ is a surprisingly good approximation. It has not seemed worth while to compute P from the full formula for $x > 3$ for any purpose. For quite small z it is seen from the graph that it is a good approximation to take $P = 1/(1+z)$ for $x > 1$, and to interpolate linearly between the points $P(0) = 1$ to $P(1) = 1/(1+z)$ on the graph for $x < 1$ —this, which will be called approximation A, leads to the simple formula

$$E' = \frac{E}{X}\left[\frac{1}{2}+\frac{X-\frac{1}{2}}{1+z}\right]. \tag{11}$$

This is the crudest approximation, which will be used later in this paper for calculations on the background count. For calculations of the actual counting rate, the results given below are to be preferred.

We can now immediately write down the expressions for the probable recorded count. For $X < 1$ the exact expressions of § 3·1 above apply; for $1 < X < 2$ we have

$$\frac{E'}{E} = \frac{1}{Xz}\{2 - e^{-zX} - e^{-z(X-1)}[1 + (X-1)z]\}$$

$$= \frac{1}{X}\left[\frac{1}{2} + \frac{X-\frac{1}{2}}{1+z} - \frac{z^2}{6}(X^3 - 6X^2 + 12X - 5) + \dots\right]. \tag{12}$$

Using $P(x) = 1/(1+z)$ for $x > 2$ the result for $X > 2$ becomes

$$\frac{E'}{E} = \frac{1}{Xz}\left[2 - e^{-2z} - e^{-z}(1+z) + \frac{(X-2)z}{1+z}\right]$$

$$= \frac{1}{X}\left[\frac{1}{2} + \frac{X-\frac{1}{2}}{1+z} - \tfrac{1}{2}z^2 + \tfrac{23}{24}z^3 - \dots\right]. \tag{13}$$

The results of equations (12) and (13) taken together are called approximation C. The less exact approximation (valid only for small z) of taking $P(x) = 1/(1+z)$ for $x > 1$, and the exact expression (equation (8)) for $x < 1$, we call approximation B. This gives, for $X > 1$,

$$\frac{E'}{E} = \frac{1}{X}\left[\frac{X-1}{1+z} + \frac{1-e^{-z}}{z}\right] = \frac{1}{X}\left[\frac{1}{2} + \frac{X-\frac{1}{2}}{1+z} - \tfrac{1}{3}z^2 + \tfrac{11}{24}z^3 - \dots\right]. \tag{14}$$

The best approximation (which is called D) uses the exact value of P up to $X = 3$, and $P = 1/(1+z)$ only thereafter. This is only required when z is of the order of unity and gives for $2 < X < 3$

$$\frac{E'}{E} = \frac{1}{Xz}[3 - e^{-Xz} - (1 + \overline{X-1}\,.\,z)\,e^{-z(X-1)} - (1 + \overline{X-2}\,.\,z + \tfrac{1}{2}\,.\,\overline{X-2}^2\,.\,z^2)\,e^{-z(X-2)}], \tag{15}$$

and for $X > 3$

$$\frac{E'}{E} = \frac{X-3}{X(1+z)} + \frac{1}{Xz}[3 - e^{-3z} - (1+2z)\,e^{-2z} - (1 + z + \tfrac{1}{2}z^2)\,e^{-z}]$$

$$= \frac{1}{X}\left[\frac{1}{2} + \frac{X-\frac{1}{2}}{1+z} - \tfrac{1}{2}z^2 + z^3 - \dots\right]. \tag{16}$$

It should be noted that this only differs from approximation C by an amount $z^3/24$ within the bracket.

Figure 4 gives graphs of E'/E for values of X up to $3\frac{1}{2}$ and z up to unity. The full curves are exact up to $X = 2$ and follow approximation C above this value. For $z = 1$ computation was carried out using approximation D; it was found that the resulting counting rate fell below that of approximation C between $X = 2{\cdot}2$ and $2{\cdot}5$, but only by $0{\cdot}3\,\%$, while as equation (16) shows, the difference decreases again as X increases beyond 3. For $z < 1$ the correction is even smaller, so that we conclude that approximation C is good enough for all practical purposes ($z > 1$ corresponds, except for quite short pulses, to conditions where more counts are lost than are recorded and is not likely to be used in practice).

The dotted curves on figure 4 show for comparison the effect of using approximation B. It is seen that this gives good values up to $z = \frac{1}{4}$, and that the error remains surprisingly small for larger values of z. We therefore recommend the use of approximation B (equation (14)) when $z < 0{\cdot}2$, and approximation C (equation (12) for $1 < X < 2$, or equation (13) for $X > 2$) for other cases, say $0{\cdot}2 < z < 0{\cdot}8$.

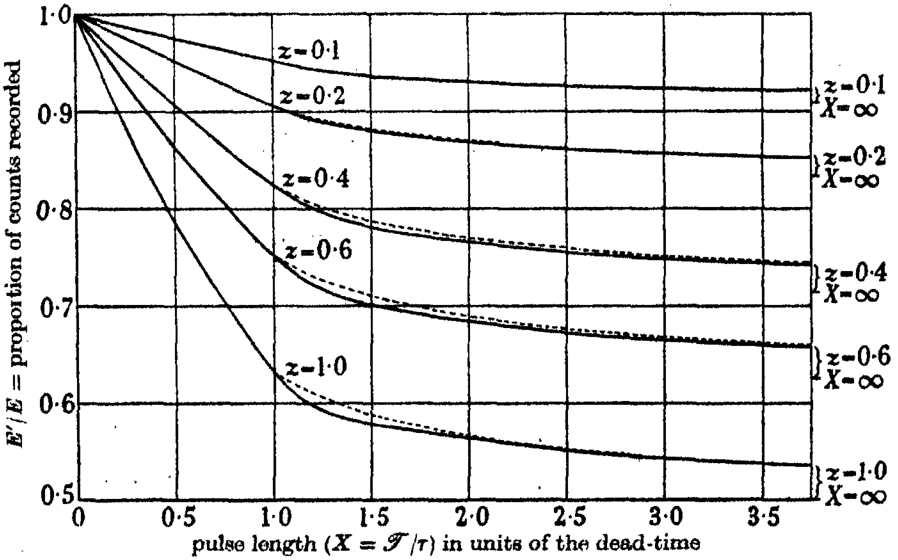


FIGURE 4. Curves giving the overall proportion of counts which are recorded. The full curves follow approximation C, and the dotted curves approximation B. Approximation D gives a curve only 0·3 % low at $z = 1{\cdot}0$, $X = 3$, and elsewhere the agreement is better. The values shown on the right as for $X = \infty$ are those which apply when counting steady (unpulsed) sources.

4. CALCULATIONS WITH A BACKGROUND COUNT

Hitherto it has been assumed that no counts at all occur in the intervals between pulses. Most practical counters have a background count, and the effect of this on the statistics must now be examined. Let μ_0 be the background count rate; then if the total true counting rate during the pulse is $\nu + \mu_0$, the total mean counting rate is $\mu = \mu_0 + f\nu = \mu_0 + \mu_1$, say. Now assume f is small, as before; and for the results obtained to have any significance, μ_1 must be at least not very much smaller than μ_0. It follows that ν/μ_0 must be quite large—in practice it is unlikely to be less than 100, and it may often be very much larger—so that terms in μ_0 can often be treated as small corrections.

4·1. *Case when* $\tau < \mathscr{T}$

Although this case, without background, has been treated fully in § 3·1 above, it is generally a condition to be avoided when low loss rates are important. With background, this case is somewhat difficult to treat, and it does not lead to a result the physical significance of which is immediately obvious. Consequently the details have been relegated to an appendix (appendix I below).

4·2. *Case when* $\tau < \mathscr{T}$

In this case both $\mu_0\tau$ and $\mu_0\mathscr{T}$ may be taken to be small, as explained above. Denoting by E'_0 the recorded count expected from background only, and by $E'_n(\mu_0+\nu)$ that predicted from equations (12) to (16) for the case without a background count but with a mean rate during the pulse of $\mu_0+\nu$ instead of ν (i.e. writing $z = (\mu_0+\nu)\tau$ in the equations), we should expect the count to be $E'_n(\mu_0+\nu)+(1-f)E'_0$, were it not that (*a*) a count in the last τ of a pulse may suppress the background for a period up to τ after the end of the pulse, and (*b*) a background count within a period τ before a pulse may suppress counts in part of the pulse.

The loss of counts from cause (*a*) may be calculated by noting that the probability of a recorded count in the last τ of any pulse is

$$\frac{(\nu+\mu_0)\tau}{1+(\nu+\mu_0)\tau} \doteqdot \frac{\nu\tau}{1+\nu\tau};$$

in obtaining this expression we take $P(t)$ as $1/(1+z)$ (cf. figure 3), since this is usually nearly true toward the end of a pulse. On the same assumption, the count is equally likely to occur at any time within the last τ of the pulse, so that on the average it will suppress the background for a period $\frac{1}{2}\tau$ after the end of the pulse, and the loss of background count will therefore be $\frac{1}{2}\mu_0\nu\tau^2/(1+\nu\tau)$ per pulse on the average.

The calculation of the loss due to (*b*) starts from the assumption that the probability of a background count within the period τ before any pulse is $\mu_0\tau$, and that this count is equally likely to occur at any time within this period. If it occurs at time $\tau-t_1$ before the pulse, it will suppress any counts in the first t_1 of the pulse; in effect, it will make the length of this pulse $\mathscr{T}-t_1$ instead of $\mathscr{T}$, for the probability function $P(t)$ will take the same form as before, but will start from time t_1 instead of 0. The net effect is therefore the same as if a period t_1 had been cut off the *end* of the pulse, and the loss of counts will therefore be $\nu t_1/(1+\nu\tau)$, taking $P(t) = 1/(1+z)$ at the end of the pulse as in considering (*a*) above. The mean value of t_1 being $\frac{1}{2}\tau$, the mean loss of counts from this cause will be $\frac{1}{2}\mu_0\nu\tau^2/(1+\nu\tau)$ per pulse. This is the same as under (*a*) above, and adding the two we obtain $\mu_0\nu\tau^2/(1+\nu\tau)$ per pulse, or $pT\mu_0\nu\tau^2/(1+\nu\tau)$ in time T. We therefore have

$$E' = E_n(\mu_0+\nu)+(1-f)E'_0-pT\mu_0\nu\tau^2/(1+\nu\tau). \tag{17}$$

5. PULSES OF NON-RECTANGULAR FORM

Throughout the foregoing the intensity of the beam of particles has been assumed to be constant throughout the pulse period $\mathscr{T}$ and zero at other times. Generally this will not be true in practice; this complicates the analysis and usually leads to integrals which can only be integrated numerically. The lines of attack to be employed in these cases can therefore only be indicated, although one case will be completely integrated below. Unfortunately, as soon as the counting speed becomes at all high ($z = \nu\tau > 0{\cdot}1$ or $0{\cdot}15$, say), a separate numerical integration is necessary for each different counting speed.

The simplest case to calculate, although the least useful in practice, occurs when the pulse length is less than the resolving time ($\mathscr{T} < \tau$); in this case not more than one count can be recorded in one pulse. It is then readily shown (see appendix II) that the probability of counting only depends on the integrated pulse intensity; writing $N(t) = \int_0^t \nu(t)\,dt$, this probability is $1 - e^{-N(\mathscr{T})}$. This method of calculation can be extended to cases where $\mathscr{T}/\tau$ does not exceed 2, 3, or even 4, but the numerical work becomes increasingly laborious. The argument is similar to that of § 3·3 above, and leads to a series in the form

$$\text{(mean count per pulse)} = \sum_{m=1}^{n} \text{(probability that an } m\text{th count occurs)}$$

terminating at the nth count, where $n-1 < \mathscr{T}/\tau \leqslant n$. The series is

$$\frac{E'}{pT} = 1 - e^{-N(\mathscr{T})} + \int_0^{\mathscr{T}-\tau} e^{-N(t)}\left(1 - e^{-N(\mathscr{T})+N(t+\tau)}\right)\nu(t)\,dt$$
$$+ \int_{t=0}^{\mathscr{T}-2\tau}\int_{s=t+\tau}^{\mathscr{T}-\tau} e^{-N(t)}\,\nu(t)\,e^{N(t+\tau)-N(s)}\,\nu(s)\left(1 - e^{N(s+\tau)-N(\mathscr{T})}\right)ds\,dt + \ldots, \qquad (18)$$

differing from the previous expressions in that $N(t+\tau)$ is a function, whereas $z(x+1)$ was a product; in addition, of course, we have $\nu(t) = (dN/dt)_t$. Numerical integration is almost always necessary, and the method is only feasible if $\mathscr{T}/\tau \leqslant 3$.

There is an alternative approach which is easier when $\mathscr{T}/\tau$ is large and the pulse envelope rises and falls fairly slowly so that $\dfrac{\tau}{\nu}\dfrac{d\nu}{dt}$ is always reasonably small. We then use the fact that the probability that a count is recorded depends on the mean counting probability during the previous τ, i.e. approximately on $\nu(t-\frac{1}{2}\tau)$, so that the probability of a recorded count in dt at t is approximately $\dfrac{\nu(t)\,dt}{1+\tau\nu(t-\frac{1}{2}\tau)}$. Each such count suppresses counts for a following τ, so that on the average it suppresses $\tau\nu(t+\frac{1}{2})$ counts. It is therefore seen that the mean number of counts suppressed during a pulse is

$$\tau\int_0^{\mathscr{T}-\frac{1}{2}\tau} \frac{\nu(t)\,\nu(t+\frac{1}{2}\tau)\,dt}{1+\nu(t-\frac{1}{2}\tau)} = \frac{E-E'}{pT}. \qquad (19)$$

It should be noted that in writing $1/(1+\tau\nu)$ for the probability of a count being recorded we are using a formula appropriate to a constant ν, and second-order errors $(O(\nu^2\tau^2))$ are to be expected; e.g. if ν is rising rapidly, conditions are similar to those of the first τ of a rectangular pulse, where we write $e^{-\nu\tau}$, which is of course $1/(1+\nu\tau+\frac{1}{2}\nu^2\tau^2+\ldots)$, while when ν is falling an error of opposite sign is expected. The integral in question terminates at $\nu(t-\tau)$ for this factor, so that the last stage of decrease is not included, and the net error is likely to have the sign corresponding to an increasing ν. Thus when we apply this integral to the rectangular pulse, we obtain the result

$$\frac{E-E'}{pT} = \frac{\nu^2(\mathscr{T}-\tau)}{1+\nu\tau} + \tfrac{1}{2}\tau\nu^2,$$

or in the notation of (12) to (16),

$$\frac{E'}{E} = \frac{1}{X}\left[\frac{1}{2} + \frac{X-\frac{1}{2}}{1+z} + \tfrac{1}{2}z^2 + \ldots\right],$$

whereas if we used $e^{-\nu\tau}$ instead of $1/(1+\nu\tau)$ for the first τ of the pulse we should obtain the $-\frac{1}{2}z^2$ term as in (13) and (16). However, when the pulse rises and falls more slowly, we shall expect the analysis of this section to be a better approximation, although we cannot carry it through exactly in the higher powers of z.

The computation of (19) would need to be repeated when the counting rate altered, but if this is not too high we can approximate, substituting $1-\nu\tau$ for $1/(1+\nu\tau)$; writing $\nu = k\nu_0$, we obtain

$$\frac{E-E'}{pT\tau} = k^2\int_0^{\mathscr{T}-\frac{1}{2}\tau} \nu_0(t)\,\nu_0(t+\tfrac{1}{2}\tau)\,dt - k^3\tau\int_{\frac{1}{2}\tau}^{\mathscr{T}-\frac{1}{2}\tau} \nu_0(t)\,\nu_0(t+\tfrac{1}{2}\tau)\,\nu(t-\tfrac{1}{2}\tau)\,dt, \tag{20}$$

so that losses can be computed for all counting speeds up to a certain limit using two integrated terms, and if we wish we may go further by continuing the expansion for a further term.

Consider as an example a pulse intensity curve of the form $\sin \pi t/\mathscr{T}$, where $\mathscr{T} = 8\tau$, so that, normalizing to $\overline{\nu_0(t)} = 1$, $\nu_0(t) = \frac{1}{2}\pi \sin \pi t/8\tau$. Remembering that $E = p\mathscr{T}kT$, we see that (20) becomes

$$\begin{aligned}\frac{E-E'}{k\tau E} &= \frac{\pi^2}{4}\int_0^{15/16} \sin\pi x\,\sin\pi(x+\tfrac{1}{16})\,dx - k\tau\frac{\pi^3}{8}\int_{1/16}^{15/16} \sin\pi x\,\sin\pi(x+\tfrac{1}{16})\,\sin\pi(x-\tfrac{1}{16})\,dx\\ &= \frac{\pi}{8}\left[\frac{15\pi}{16}\cos\frac{\pi}{16}+\sin\frac{\pi}{16}-k\tau\left\{\frac{\pi}{3}\cos\frac{3\pi}{16}+\cos\frac{\pi}{16}\right\}\right],\end{aligned}$$

or

$$(E-E')/E = 1{\cdot}212k\tau - 1{\cdot}552(k\tau)^2.$$

Comparing this with a rectangular pulse of length $X\tau$, noting that to maintain the same true mean counting rate we require $z = \nu\tau = 8k\tau/X$, and using (13), we see that

$$\begin{aligned}\frac{E-E'}{E} &= \left(1-\frac{1}{2X}\right)\left(1-\frac{1}{1+z}\right)+\frac{1}{2X}z^2-\dots \quad \text{must correspond to}\\ &= 1{\cdot}212Xz/8 - 1{\cdot}552(Xz/8)^2.\end{aligned}$$

To make the first terms agree, we must put $X = 6{\cdot}06$, i.e. take the rectangular pulse length $75\frac{3}{4}$ % of the length of the semi-sinusoidal pulse; the second term calculated for the rectangular pulse then becomes $-1{\cdot}455(Xz/8)^2$, which agrees surprisingly well with the $1{\cdot}552$ term obtained with a pulse of different form. The first-order result, that a semi-sinusoidal pulse behaves like a square pulse of 75 % of its length, is also clearly reasonable (figure 5).[*]

[*] For $\tau \ll \mathscr{T}$ equation (20) simplifies. We obtain

$$\frac{E-E'}{E} = \frac{k\tau}{\mathscr{T}}\int_0^{\mathscr{T}} [\nu_0(t)]^2\,dt - \frac{k^2\tau^2}{\mathscr{T}}\int_0^{\mathscr{T}} [\nu_0(t)]^3\,dt + \dots,$$

while for a rectangular pulse of length $\mathscr{T}_{\text{eff.}}$, say, giving the same total count, we must have $z = k\tau\mathscr{T}/\mathscr{T}_{\text{eff.}}$, since $\overline{\nu_0(t)} = 1$, and

$$\frac{E-E'}{E} = 1-\frac{1}{1+z} = z-z^2+\dots.$$

For the two pulses to be equivalent, we may equate coefficients; from the first term we obtain

$$\frac{\mathscr{T}}{\mathscr{T}_{\text{eff.}}} = \int_0^{\mathscr{T}} [\nu_0(t)]^2\,dt \quad \text{or} \quad \mathscr{T}_{\text{eff.}} = \frac{\left[\int_0^{\mathscr{T}} \nu(t)\,dt\right]^2}{\int_0^{\mathscr{T}} [\nu(t)]^2\,dt}.$$

Before leaving this matter of pulse form, there is another important factor which must not be overlooked. Most particle accelerators are such that they give, instead of a single pulse of particles, a number of pulses at regular intervals, usually corresponding to a radio-frequency used in acceleration. This fine structure of the pulse usually passes unnoticed, but unless its period is considerably smaller than τ it will affect the distribution, and therefore the losses, of the resulting counts. The radio-frequencies used are usually 10 Mc./sec. or higher, so that unless $\tau < 5\cdot10^{-7}$ sec. (not very easy to achieve with present techniques), the effect will be negligible. It should be remembered, however, that should slower fine-structure effects occur, e.g. as a result of precession of the orbits of the particles in the accelerator, their effect on the counting losses may be appreciable. Moreover, as counting techniques giving higher resolution are introduced, effects of this kind will increase in importance. The magnitude of the effect can clearly be estimated by applying the methods developed in this paper for the pulse as a whole, although the additional complication introduced may be considerable.

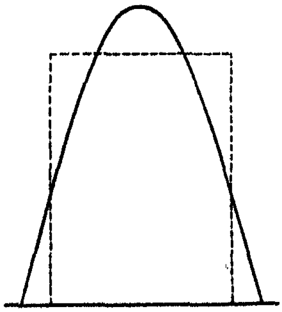


FIGURE 5. Illustrating the effect of varying the pulse envelope. The semi-sinusoidal pulse gives the same total counting rate as the rectangular one shown dotted, as well as the same first-order counting loss probability. The second-order terms differ only slightly. The height of the rectangle is 84 % of the sine-wave peak and its width $75\frac{3}{4}$ % ($\mathscr{T} = 8\tau$). For $\mathscr{T} = 4\tau$, similar calculations give a width of 72 % and a height of 88·3 %, while, for $\tau \ll \mathscr{T}$, the considerations of the footnote lead to a width of 81·5 % and a height of 78·5 %.

6. CONCLUSION

Although it has been possible only to indicate methods to be followed in the case of the pulse of general shape, it is believed that in most practical cases the applications of these methods will not be unduly difficult. Frequently, first order corrections will suffice, though it may be desirable to satisfy oneself, by working one case, that higher order terms are negligible. For example, with $p = 100\,\text{sec.}^{-1}$, $\mathscr{T} = 5\,\mu\text{sec.}$, $\tau = 2\,\mu\text{sec.}$, and a mean counting rate of 3000 per min., we have $X = 2\cdot5$ and $z = 0\cdot2$. The first-order term is then $E'/E = 0\cdot4(\frac{1}{2} + 2/1\cdot2) = 13/15$, giving about a 13 % counting loss; the second order term in (13) is only $-0\cdot8$ %. These are typical of the conditions one may meet in practice.

It is, of course, true for all these calculations that we have assumed the true counting rate and deduced the recorded one, whereas the actual problem is the converse. In practice, however, this causes little difficulty, a process of successive approximations leading very rapidly to the true result; this must be so unless the counting loss is very high, of the order of 50 % or more, and such conditions should never be used when results of any accuracy are required.

A further point which deserves mention is that conditions are even more stringent than those we have considered when the machines use a pulsed extraction system, with which the output is confined to one radio-frequency cycle. This is commonly done with the largest cyclotrons, and for synchrotrons giving energies of the order of 10^9 eV is almost the only method of extraction so far considered. In these cases $\mathscr{T}$ becomes of the order of 10^{-8} sec., so that with the counting techniques at present available we can never attain the desirable condition $\tau < \mathscr{T}$. The moral to be drawn is that pulsed extraction is to be avoided whenever possible, the only alternative if reasonably fast counting is desired being the development of extremely high resolution counting arrangements, possibly using crystal counters.

It should also be made clear that any consideration of the effect of using a pulsed source with coincidence counting arrangements has been excluded from the present paper. One may, for example, wish to use coincidence technique to reduce the background count due to neutron recoils, etc., or to identify particles of a given range, as well as for observing genuine multiple effects. That additional difficulties arise in this case when a pulsed source is used has been mentioned by Wilson, Lofgren, Richardson, Wright & Shankland (1947) in a brief note, although they did not give any calculations on the subject. A full discussion would be quite lengthy, and cannot be included here.

The author wishes to express his thanks to Professor M. L. E. Oliphant for his interest and helpful advice, and to acknowledge assistance from several colleagues who read this paper in draft and made suggestions.

APPENDIX I. CALCULATION WITH BACKGROUND COUNTS WHEN $\tau > \mathscr{T}$

(N.B. The term 'pulse interval' is used for the period $1/p - \mathscr{T}$ between the end of one pulse and the beginning of the next.)

In this case we may assume (see § 4 above) that $\mu_0\mathscr{T}$ is small, although $\mu_0\tau$ may not be. We shall, however, assume that the pulse interval is long enough for the mean rate of recording the background counts to have settled down after the disturbance due to a pulse, for at least a period τ before the next pulse, to its steady value of $\mu_0/(1+\mu_0\tau)$. As will be seen below, this requires that $1/p - \mathscr{T} > 2\frac{1}{2}\tau$. We may calculate the probability of a count occurring during the pulse as in § 3·1 above, remembering that there cannot be more than one such count. If no recorded background count has occurred in the τ preceding the pulse (probability $1/(1+\mu_0\tau)$*), the probability of counting during the pulse is just as in § 3·1 above. If a recorded background count has occurred within $\tau - \mathscr{T}$ before the pulse, the probability of counting during the pulse is zero, and if it has occurred at a time $\tau - t'$ before the pulse ($t' < \mathscr{T}$), the probability is given by reducing the pulse length in (3) from $\mathscr{T}$ to $\mathscr{T} - t'$. We have, therefore, for the total probability of counting during the pulse,

---

* Not $e^{-\mu_0\tau/(1+\mu_0\tau)}$, since only one recorded count can occur in any τ.

$$\begin{aligned}\Pi &= \frac{1}{1+\mu_0\tau}(1-e^{-(\nu+\mu_0)\mathscr{T}}) + \int_0^{\mathscr{T}} \frac{\mu_0\,dt'}{1+\mu_0\tau}(1-e^{-(\nu+\mu_0)(\mathscr{T}-t')}) \\ &= \frac{1}{1+\mu_0\tau}\left[\mu_0\mathscr{T} + 1 - \frac{\nu}{\nu+\mu_0}e^{-(\nu+\mu_0)\mathscr{T}}\right]. \qquad (21)\end{aligned}$$

We see that the first term represents Π_0, the probability of counting during this period if only the background were present, so that we may separate

$$\begin{aligned}\Pi - \Pi_0 &= \{1 - \nu e^{-(\nu+\mu_0)\mathscr{T}}/(\nu+\mu_0)\}/(1+\mu_0\tau) \\ &= \text{the probability of a count not due to background occurring}\end{aligned}$$

from $1 - \Pi + \Pi_0 =$ the probability that conditions during the pulse are just as they would have been if no pulse had occurred. In the latter case the rate of occurrence of recorded counts throughout the interval following the pulse in question remains equal to $\mu_0/(1+\mu_0\tau)$, so that the expected recorded count during this interval is $\mu_0(1/p - \mathscr{T})/(1+\mu_0\tau)$.

In the other case, probability $\Pi - \Pi_0$, we know that a count has occurred in the pulse, so for a period of between $\tau - \mathscr{T}$ and τ afterwards (the average value $\tau - \frac{1}{2}\mathscr{T}$ can be used for calculations) no background count is possible. The first true background count after the end of this period is necessarily recorded, and then things proceed normally; consequently, on the average, after a pulse giving a recorded count, the system is sensitive to background for a period

$$1/p - \mathscr{T} - (\tau - \tfrac{1}{2}\mathscr{T}) = 1/p - \tfrac{1}{2}\mathscr{T} - \tau.$$

The mean rate of recording counts during this period is $\mu_0/(1+\mu_0\tau)$, except that at the beginning we know that there has been a dead period, so that the first true count will certainly record and the probable rate must be μ_0. Using approximation A (cf. (11)) we take the rate to be μ_0 for the first $\frac{1}{2}\tau$ and $\mu_0/(1+\mu_0\tau)$ thereafter.*

We then obtain for the probable count in a pulse interval

$$\begin{aligned}&(1-\Pi+\Pi_0)(1/p-\mathscr{T})\frac{\mu_0}{1+\mu_0\tau} + (\Pi-\Pi_0)\left\{\tfrac{1}{2}\mu_0\tau + \left(\frac{1}{p} - \tfrac{1}{2}\mathscr{T} - \tfrac{3}{2}\tau\right)\frac{\mu_0}{1+\mu_0\tau}\right\} \\ &\qquad = \frac{\mu_0}{1+\mu_0\tau}\left[\frac{1}{p} - \mathscr{T} - (\Pi-\Pi_0)(\tau - \tfrac{1}{2}\mu_0\tau^2 - \tfrac{1}{2}\mathscr{T})\right]. \qquad (22)\end{aligned}$$

Substituting for $\Pi - \Pi_0$ and adding to Π, we obtain a value for the mean counting rate of

$$\frac{p}{(1+\mu_0\tau)^2}\left[1 - \frac{\nu}{\nu+\mu_0}e^{-(\nu+\mu_0)\mathscr{T}}\right]\{1 + \tfrac{1}{2}\mu_0(\mathscr{T} + \mu_0\tau^2)\} + \frac{\mu_0}{1+\mu_0\tau}, \qquad (23)$$

from which we see that, since $E_0 = \mu_0 T$ and $E_0' = \mu_0 T/(1+\mu_0\tau)$,

$$\frac{E' - E_0'}{E - E_0} = \frac{1}{\nu\mathscr{T}(1+\mu_0\tau)^2}\{1 + \tfrac{1}{2}\mu_0(\mathscr{T} + \mu_0\tau^2)\}\left[1 - \frac{\nu}{\nu+\mu_0}e^{-(\nu+\mu_0)\mathscr{T}}\right], \qquad (24)$$

which is readily evaluated numerically. To the first order in μ_0 the right-hand side is

$$\frac{1-e^{-\nu\mathscr{T}}}{\nu\mathscr{T}}\{1 - \mu_0(2\tau - \tfrac{1}{2}\mathscr{T})\} + \frac{\mu_0}{\nu}\left(1 + \frac{1}{\nu\mathscr{T}}\right)e^{-\nu\mathscr{T}}. \qquad (24a)$$

---

* Note that this procedure leads to the figure of $2\frac{1}{2}\tau$ given at the end of the first paragraph of this appendix.

APPENDIX II. NON-UNIFORM COUNTING RATE RESULTS

Let $\nu(t)\,dt$ be the probability of a count occurring in dt at t, and $N = \int_0^t \nu(t)\,dt$. Let $P(0,t)$, $P(1,t)$, etc. be the probability that 0, 1, etc. counts have occurred between $t = 0$ and t. Then $P(0,0) = 1$, $P(1,0)$ etc. $= 0$. Also

$$P(m.t+dt) = P(m,t)\,(1-\nu(t)\,dt) + P(m-1,t)\,\nu(t)\,dt,$$

$$\frac{d}{dt}(P(m,t)) = \nu(t)\,(P(m-1,t)-P(m,t)).$$

But $\nu(t) = \dfrac{dN}{dt}$ by definition; therefore $\dfrac{d}{dN}(P(m,t)) = P(m-1.t)-P(m.t)$.

This is the same relation that would have been obtained had ν been constant and $N = \nu t$, so that the P's are the same function of N as in the constant rate case. Therefore, as may also be proved directly,

$$P(0,t) = e^{-N}, \quad P(1,t) = Ne^{-N} \quad \text{and generally} \quad P(n,t) = N^n e^{-N}/n!.$$

The probability of the $(n+1)$th count occurring in dt at t is clearly $P(n,t)\,\nu(t)\,dt$.

APPENDIX III. SUMMARY OF RESULTS

(*For list of symbols see p.* 508)

This appendix is included for rapid reference, as well as for the experimenter requiring the results without wishing to follow the analysis in detail.

(1) *Uniform (unpulsed) counting*

$$\frac{E'}{E} = \frac{1}{1+\mu\tau}. \tag{2}$$

(2) *Rectangular pulse form, no background*

(*a*) $\mathscr{T}/\tau < 1$:

$$\frac{E'}{E} = \frac{1-e^{-\nu\mathscr{T}}}{\nu\mathscr{T}} \quad \text{(from (3))}.$$

(*b*) $\mathscr{T}/\tau \gg 1$:

$$\frac{E'}{E} = 1-\mu\left(1-\frac{\tau}{2\mathscr{T}}\right)\frac{\tau}{f} \quad \text{(from (5))}.$$

(*c*) $\mathscr{T}/\tau > 1$ but not $\gg 1$ (N.B. $\mathscr{T}/\tau = X$).

(i) $z < 0{\cdot}2$ (N.B. $z = \nu\tau$):

$$\frac{E'}{E} = \frac{1}{X}\left[\frac{X-1}{1+z}+\frac{1-e^{-z}}{z}\right] = \frac{1}{X}\left[\frac{1}{2}+\frac{X-\frac{1}{2}}{1+z}-\frac{z^2}{3}+\dots\right]. \tag{14}$$

(ii) $z > 0{\cdot}2$, $1 < \mathscr{T}/\tau < 2$:

$$\frac{E'}{E} = \frac{1}{Xz}\{2-e^{-zX}-e^{-z(X-1)}[1+(X-1)z]\} \quad \text{(or expanded as in (12))}.$$

(iii) $0{\cdot}2 < z < 0{\cdot}8$, $\mathscr{T}/\tau > 2$:

$$\frac{E'}{E} = \frac{1}{Xz}\left\{2 - e^{-2z} - e^{-z}(1+z) + \frac{(X-2)\,z}{1+z}\right\} \quad \text{(or expanded as in (13)).}$$

(iv) $z > 0{\cdot}8$; see equations (15) and (16) for $X > 2$. (This highest approximation is seldom required.)

(3) *Non-rectangular pulses, no background*

(*a*) $\mathscr{T}/\tau < 1$. Writing $N(\mathscr{T}) = \int_0^{\mathscr{T}} \nu(t)\,dt$, we obtain

$$\frac{E'}{E} = \frac{1 - e^{-N(\mathscr{T})}}{N(\mathscr{T})}.$$

(*b*) $1 < \mathscr{T}/\tau \leqslant 3$. Use (18), the left-hand side of which is $N(\mathscr{T})\,E'/E$.

(*c*) $\mathscr{T}/\tau$ fairly large $\left(\text{i.e., } \frac{\tau}{\nu}\frac{\partial \nu}{\partial t} \leqslant \frac{1}{3}\right)$.

$$1 - \frac{E'}{E} = \tau N(\mathscr{T}) \int_0^{\mathscr{T}-\frac{1}{2}\tau} \frac{\nu(t)\,\nu(t+\frac{1}{2}\tau)}{1+\nu(t-\frac{1}{2}\tau)}\,dt \quad \text{(from (19)).}$$

This is valid only for $\nu\tau$ not too large; the integral has to be re-evaluated for every change of counting rate, but this may be avoided by expanding, as in (20). A useful procedure is to calculate the first-order counting loss from (20) and thence to deduce the equivalent length of a rectangular pulse giving this first-order loss, subsequently using the equations of § 2 of this appendix.

(4) *Rectangular pulses with background present*

(*a*) $\mathscr{T}/\tau < 1$. Use (24), appendix I. The expressions simplify when the background rate is very low so that losses therein can be neglected ($\mu_0\tau$ and μ_0/ν negligible); we then obtain (cf. § 2 (*a*) of this appendix)

$$\frac{E' - E'_0}{E - E_0} = \frac{1 - e^{-\nu\mathscr{T}}}{\nu\mathscr{T}}.$$

(*b*) $\mathscr{T}/\tau > 1$:

$$E' - E'_n(\mu_0 + \nu) = (1-f)\,E'_0 - pT\mu_0\nu\tau^2/(1+\nu\tau) \quad \text{(from (17)),}$$

which reduces the problem to calculating $E'_n(\mu_0+\nu)$, i.e. to a case without background.

REFERENCES

Bateman, H. 1910 *Phil. Mag.* (6), **20**, 704.
Hole, N. 1946 *Ark. Mat. Astr. Fys.* **33** A, no. 11.
Locher, G. L. 1933 *J. Franklin Inst.* **216**, 553.
Marsden, E. & Barratt, T. 1911 *Proc. Phys. Soc. Lond.* **23**. 367
Ruark, A. E. & Brammer, F. E. 1937 *Phys. Rev.* **52**, 322.
Skinner, S. M. 1935 *Phys. Rev.* **48**, 438.
Wilson, R. R, Lofgren, E. J., Richardson, J. R., Wright, B. T. & Shankland, R. S. 1947 *Phys. Rev.* **71**, 560.
Volz, H. 1935 *Z. Phys.* **93**, 539.

Decay of turbulence in the final period

By G. K. Batchelor, *Trinity College, University of Cambridge*
and A. A. Townsend, *Emmanuel College, University of Cambridge*

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 28 *January* 1948)

The final period of decay of a turbulent motion occurs when the effects of inertia forces are negligible. Under these conditions the instantaneous velocity distribution in the turbulence field may be solved as an initial value problem. It is shown that homogeneous turbulence tends to an asymptotic statistical state which is independent of the initial conditions. In this asymptotic state the energy of turbulence is proportional to $t^{-\frac{5}{2}}$ and the longitudinal double-velocity correlation coefficient for two points distance r apart is $e^{-r^2/8\nu t}$, where t is the time of decay. The asymptotic time-interval correlation coefficient is found to be different from unity for very large time intervals only, showing the aperiodic character of the motion. The whole field of motion comes gradually to rest, smaller eddies decaying more rapidly than larger eddies, and the above stable eddy distribution is established when only the largest eddies of the original turbulence remain.

Relevant measurements have been made in the field of isotropic turbulence downstream from a grid of small mesh. The above energy decay and space-interval correlation relations are found to be valid at distances from the grid greater than 400-mesh lengths and at a mesh Reynolds number of 650. The duration of the transitional period, in which the energy decay law is changing from that appropriate to the initial period of decay to the above asymptotic law, increases very rapidly with R_M. There is a brief discussion of the criterion for the existence of final period decay, although clarification must wait until the existence and termination of the initial period of decay are better understood.

1. Introduction

It has long been a standard practice in hydrodynamical research to seek conditions under which the Navier-Stokes equations become linear. For instance, the theories of slow motion of a viscous fluid, and of small perturbations to a basic flow, have each been well developed. It is thus natural to inquire if there are any circumstances under which the quadratic inertia terms play no part in the problem of turbulent motion. The immediate answer is that when the Reynolds number is low enough the inertia forces are negligible compared with the viscous forces, and the governing equations are effectively linear. Despite this tremendous gain in simplicity, considerations of low Reynolds states of turbulence are rare in the literature, and not always accurate. Theoretical investigators may have hesitated on account of the fact that no experimental evidence demonstrating clearly that inertia forces were negligible has yet been presented. It will be seen below that this is simply because measurements of turbulence have not been made at sufficiently low Reynolds numbers.

It is the essence of the phenomenon of turbulence that there is a continual transfer of energy from large to small eddies. This transfer, which maintains the high rate of dissipation characteristic of turbulent motion, is entirely a result of the action of inertia effects. Consequently it is slightly anomalous to speak of a low Reynolds number state of turbulent motion in which inertia effects play no part; local regions of the field will in fact have a stable *laminar* motion. The motion is only turbulent in the sense that it is the last stage of decay of a motion which was formerly turbulent in the precise sense. But it would be too confusing to term it a stable or laminar state of turbulence, and we propose that it be called the final period of decay of the turbulence. The term has already been suggested in a previous paper (Batchelor & Townsend 1948) in a consideration of the decay of isotropic turbulence. In that paper it was shown that there is an initial period in the decay of isotropic turbulence during which the energy decays according to the simple law

$$\overline{u^2} \propto t^{-1}, \tag{1·1}$$

and the viscous and inertia forces are of comparable importance, and that after a certain time there is a transitional period during which the law of energy decay is changing. The inference—it was not possible to obtain direct evidence by making measurements at greater times of decay—was that the law of energy decay was tending towards that which ultimately obtains in the final period, viz.

$$\overline{u^2} \propto t^{-\frac{5}{2}}, \tag{1·2}$$

the inertia forces then being negligible. The measurements to be presented below are concerned wholly with isotropic turbulence and do in fact confirm the law (1·2). But it should be noted that all decaying turbulent motions will have their final period, with certain characteristics in common. Some of these common features emerge from the analytical discussion to be given in §§ 3 and 4.

2. Recapitulation of previous discussions of isotropic turbulence at small Reynolds numbers

All previous investigators have considered the law of energy decay and the double-velocity correlation. The turbulent energy per unit mass of fluid is $\frac{3}{2}\overline{u^2}$, while the double-velocity correlation is completely specified by $\overline{u^2}$ and the function $f(r,t)$, where $\overline{u^2}f(r,t)$ is the correlation at time t between parallel velocity components at two points distance r apart along a line parallel to the velocities. With the neglect of any inertia forces, the Navier-Stokes equations supply the following equation for f and $\overline{u^2}$:

$$\frac{\partial f\overline{u^2}}{\partial t} = 2\nu\left(\frac{\partial^2 f\overline{u^2}}{\partial r^2} + \frac{4}{r}\frac{\partial f\overline{u^2}}{\partial r}\right), \tag{2·1}$$

where ν is the kinematic viscosity of the fluid. When $r = 0$, (2·1) degenerates to the energy equation

$$\frac{d\overline{u^2}}{dt} = 10\nu\overline{u^2}f_0'' = -\frac{10\nu\overline{u^2}}{\lambda^2}, \tag{2·2}$$

where λ is the dissipation length parameter.

Von Kármán & Howarth (1938) appear to be the first authors to have neglected deliberately the inertia effects in order to investigate isotropic turbulence at low Reynolds numbers. They considered only those solutions of (2·1) which are functions of r/λ ($= \psi$, say) alone, and effectively assumed that the energy decays according to the law

$$\overline{u^2} \propto t^{-n}, \tag{2·3}$$

and thence that $\lambda^2 = (10/n)\,\nu t$, where n is left undetermined. They showed that there is a singly infinite family of such solutions involving n as parameter, viz.

$$f(\psi) = \left(\frac{4n}{5}\right)^{\frac{5}{4}} \psi^{-\frac{5}{2}} \exp\left[-\frac{5}{8n}\psi^2\right] M_{n-\frac{5}{4},\frac{3}{4}}\left(\frac{5}{4n}\psi^2\right), \tag{2·4}$$

where $M_{k,m}(z)$ is a confluent hypergeometric function. There does not appear to be any absolute reason why the solution (2·3) and (2·4) for an arbitrary value of n should not represent the way in which the turbulence changes in practice provided that the appropriate correlation function can be produced at any instant. But as has already been pointed out (Batchelor 1948), it appears to be possible to do this only when $n = \frac{5}{2}$, so far as turbulence produced in the laboratory is concerned. When $n < \frac{5}{2}$, the function (2·4) is such that $\int_0^\infty \psi^4 f(\psi)\,d\psi$ is infinite, which is unlikely to be true under practical conditions, while when $n > \frac{5}{2}$, $\int_0^\infty \psi^4 f(\psi)\,d\psi = 0$ and $f(\psi)$ is negative for a certain range of values of ψ, unlike measured values of f. Since Λ, $= \overline{u^2}\int_0^\infty r^4 f(r)\,dr$, is independent of the time of decay (Loitsiansky 1939), if $\int_0^\infty r^4 f\,dr$ is neither zero not infinite at any instant, it will remain so. Thus the particular case $n = \frac{5}{2}$, which leads to

$$\overline{u^2} \propto t^{-\frac{5}{2}}, \quad f(r,t) = e^{-r^2/8\nu t}, \tag{2·5}$$

is the self-preserving solution which one might expect to find in an experiment.

In the same year, Reissner (1938) considered initial-value solutions of the equation

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2} \tag{2·6}$$

for the velocity $u(x,t)$ in a one-dimensional turbulence field, with the notion that this represents a simplified model of real turbulent motion. He finds that the correlation function at time t is related to the velocity field at time t_0 by

$$\overline{u^2} f(r,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-2\nu\lambda^2(t-t_0)} \cos\lambda r \lim_{\alpha\to\infty} \frac{\left[\int_{-\alpha}^{\alpha} u(x,t_0)\cos\lambda x\,dx\right]^2}{2\alpha}\,d\lambda. \tag{2·7}$$

Letting $t\to\infty$ to obtain asymptotic expressions (which is a legitimate and useful operation in an inquiry into conditions in the final period of decay), Reissner shows that(2·7) becomes

$$\overline{u^2} \to \frac{1}{\sqrt{[8\pi\nu(t-t_0)]}} \lim_{\alpha\to\infty} \frac{\left[\int_{-\alpha}^{\alpha} u(x,t_0)\,dx\right]^2}{2\alpha}, \tag{2·8}$$

the correlation coefficient f being effectively unity. He states an analogous result for three-dimensional turbulence in which each velocity component obeys the equation

$$\frac{\partial u}{\partial t} = \nu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right), \tag{2·9}$$

viz.

$$\overline{u^2} \to \frac{\text{const.}}{(t-t_0)^{\frac{3}{2}}} \lim_{\alpha,\beta,\gamma\to\infty} \frac{\left[\int_{-\alpha}^{\alpha}\int_{-\beta}^{\beta}\int_{-\gamma}^{\gamma} u(x,y,z,t_0)\,dx\,dy\,dz\right]^2}{8\alpha\beta\gamma}. \tag{2·10}$$

This energy decay law of Reissner is not correct, even though equation (2·9) represents correctly the motion of the fluid in the final period. It can be shown that as a consequence of the condition of incompressibility of the fluid, the constant in (2·10) is identically zero, and a higher order term varying as $(t-t_0)^{-\frac{5}{2}}$ replaces the expression (2·10). It will be seen in the following sections that Reissner's idea of solving the equation for the instantaneous local velocity in the field is capable of yielding many useful results.

Loitsiansky (1939) has remarked that equation (2·1) is similar to the equation for the conduction of heat in a spherically symmetrical five-dimensional temperature field. Using this analogy, the relation between $\overline{u^2}f$ at times t and t_0 is known to be

$$\overline{u^2}(t)f(r,t) = \frac{1}{[8\pi\nu(t-t_0)]^{\frac{5}{2}}}\iiiint\!\!\!\int_{-\infty}^{\infty}\overline{u^2}(t_0)f(s,t_0)\exp\left[-\frac{\rho^2}{8\nu(t-t_0)}\right]dx_1\,dx_2\,dx_3\,dx_4\,dx_5, \tag{2·11}$$

where $$\rho^2 = \sum_{n=1}^{5}(\xi_n-x_n)^2, \quad r^2 = \sum_1^5 \xi_n^2, \quad s^2 = \sum_1^5 x_n^2.$$

The law of energy decay is found by putting $r = 0$ in (2·11), i.e.

$$\overline{u^2}(t) = \frac{8\pi^2\overline{u^2}(t_0)}{3[8\pi\nu(t-t_0)]^{\frac{5}{2}}}\int_0^\infty f(s,t_0)\exp\left[-\frac{s^2}{8\nu(t-t_0)}\right]s^4\,ds. \tag{2·12}$$

Thus, in general, the law of energy decay and the correlation function depend on, and are uniquely determined by, the conditions at time t_0. Loitsiansky points out that for the special initial conditions given by

$$\left.\begin{aligned} (\overline{u^2}f)_{t_0} &= \infty \quad \text{when} \quad r = 0,\\ &= 0 \quad \text{when} \quad r > 0,\\ \int_0^\infty (\overline{u^2}f)_{t_0}r^4\,dr &= \Lambda \text{ (finite)}, \end{aligned}\right\} \tag{2·13}$$

the integrals of (2·11) and (2·12) can be evaluated to give

$$\overline{u^2}(t) = \frac{\Lambda}{48\sqrt{(2\pi)}}[\nu(t-t_0)]^{-\frac{5}{2}}, \tag{2·14}$$

$$f(r,t) = \exp\left[-\frac{r^2}{8\nu(t-t_0)}\right]. \tag{2·15}$$

These special initial conditions are described as referring to a 'point source of strength Λ' in view of the corresponding problem in the heat analogy.

Millionshtchikov (1939) independently obtained this special solution in the same way; he interpreted the solution as describing the turbulence which exists subsequent to an initial random distribution of concentrated line eddies, provided that the effect of triple correlations is ignored. This is true in the sense that initially the velocity correlation is non-zero for $r = 0$ only, but the interpretation does not account for the third (and most important) of the initial conditions (2·13). In any case it is not practically possible to create a state of turbulence in which inertia forces are negligible during its whole history, so that the special initial conditions used by Loitsiansky and Millionshtchikov have no reference to real turbulence.

Sedov (1944) covered much the same ground as von Kármán & Howarth, Loitsiansky, and Millionshtchikov, and pointed out the relation between the parameter n in von Kármán & Howarth's family of solutions and the value of $\overline{u^2}\int_0^\infty r^4 f\,dr$.

Finally, it was shown by one of us (Batchelor 1948) that, whatever the conditions at $t = t_0$, the energy decay and correlation function derived from equation (2·1) tend to the forms (2·14) and (2·15) as $t \to \infty$. This follows readily from a consideration of the general solutions (2·11) and (2·12) at large values of $t - t_0$ (provided that we accept the restrictions that Λ is neither zero nor infinite). There is thus a unique asymptotic solution in the final period, the asymptotic correlation function being completely self-preserving in shape. It was also shown that such a self-preserving correlation function can *only* exist when inertia forces are negligible, and is necessarily an asymptotic solution under such conditions.

Mathematically, the position is this: that of the infinite family of self-preserving solutions derived by von Kármán & Howarth, a certain number only are asymptotic solutions for arbitrary initial conditions. Initial conditions for which $\int_0^\infty r^4 f\,dr$ is neither zero nor infinite (as will normally be the case in practice) give an asymptotic decay law such that $\overline{u^2} \propto t^{-\frac{5}{2}}$, corresponding to $n = \frac{5}{2}$ in the family of solutions. If the initial conditions are such that $\int_0^\infty r^4 f\,dr = 0$ but $\int_0^\infty r^6 f\,dr \neq 0$, the asymptotic decay law is obtained by taking the next term in the expansion of the exponential in (2·12) and is of the form $\overline{u^2} \propto t^{-\frac{7}{2}}$, corresponding to $n = \frac{7}{2}$. Likewise, if $\int_0^\infty r^4 f\,dr$ is infinite but $\int_0^\infty r^2 f\,dr$ is finite, the asymptotic decay law is $\overline{u^2} \propto t^{-\frac{3}{2}}$, corresponding to $n = \frac{3}{2}$; and so on, all half-integral values of n providing possible asymptotic solutions. The corresponding asymptotic correlation functions are readily obtained from either (2·4) or (2·11).

Physically, the conditions under which turbulence is produced (at any rate, in a laboratory) will ensure that $\int_0^\infty r^4 f\,dr$ is neither zero nor infinite. Hence the solution which will be found when the inertia forces are negligible, and when the time of decay is sufficiently large, is

$$\overline{u^2} \propto t^{-\frac{5}{2}}, \quad f(r,t) = e^{-r^2/8\nu t}. \tag{2·16}$$

3. NEW APPROACH TO THE PROBLEM

A new approach which, like Reissner's investigation, considers the local or unaveraged turbulent velocities will now be presented. The asymptotic decay law and correlation function already obtained, and other new results of interest, will be deduced ab initio. We confine attention to the case of zero mean motion.

When the Reynolds number of the turbulent motion is so low that inertia forces are negligible, the Navier-Stokes equation becomes

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + \nu\nabla^2 u, \tag{3·1}$$

where u is a typical velocity component. Using the equation of continuity, the pressure is found to satisfy the equation

$$\nabla^2 p = 0.$$

Since a potential function cannot have a maximum or a minimum in the interior of the field, the only possible bounded solution for an arbitrarily large fluid field not containing solid boundaries is that p is independent of position. Hence the equation for u reduces to

$$\frac{\partial u}{\partial t} = \nu \nabla^2 u. \tag{3·2}$$

Each velocity component therefore changes with time in the final period of decay in the same way as a similar distribution of temperature in a homogeneous medium. The velocity components change independently of each other, and if the equation of continuity is satisfied at any time, it remains satisfied at all subsequent times. Since the velocity components do not interact with one another, symmetry conditions on the turbulence become of secondary importance and any component can be considered in isolation.

If the spatial distribution of u is known at the instant from which the time is measured, the solution of (3·2) is

$$u(x,y,z,t) = \frac{1}{(4\pi\nu t)^{\frac{3}{2}}} \int_{-\alpha}^{\alpha} \int_{-\beta}^{\beta} \int_{-\gamma}^{\gamma} u(X,Y,Z,0) \times \exp\left[-\frac{(X-x)^2-(Y-y)^2-(Z-z)^2}{4\nu t} \right] dX\,dY\,dZ, \tag{3·3}$$

as is well known from the study of other physical problems governed by the equation (3·2). The field of turbulence is here assumed to have (large) dimensions 2α, 2β and 2γ in the x, y and z directions. The restriction to a finite field is not significant and could be removed—at the expense of simplicity—by using Fourier transforms of the velocity distribution in a manner similar to Reissner's investigation.

The turbulence can thus be calculated if it is known at some instant. Naturally, only the statistical characteristics are likely to be known at the initial instant, so that (3·3) is not directly useful in its present form. The task is to use it to deduce the development with time of various mean values. The advantage of this explicit solution for the velocity field is that it provides a means (which may not always be practical!) of determining the development of *any* of the statistical characteristics of turbulence during the final period of decay.

The mean value of any quantity associated with the turbulence will be regarded as defined by a space integral, viz.

$$\bar{Q}(t) = \frac{1}{8\alpha\beta\gamma} \int_{-\alpha}^{\alpha} \int_{-\beta}^{\beta} \int_{-\gamma}^{\gamma} Q(x,y,z,t)\,dx\,dy\,dz. \tag{3·4}$$

Implicit in this definition of a mean value is the assumption that the turbulence is approximately homogeneous.

4. DOUBLE VELOCITY CORRELATIONS IN THE FINAL PERIOD

Consider first the correlation between two parallel simultaneous velocity components at two points P and P', where PP' has the projection ξ in the direction of the velocity components and projections η and ζ in two other orthogonal directions. Using (3·4),

$$\begin{aligned} R(\xi,\eta,\zeta,t) &= \overline{u(x,y,z,t)\,u(x+\xi,y+\eta,z+\zeta,t)} \\ &= \frac{1}{8\alpha\beta\gamma}\int_{-\alpha}^{\alpha}\int_{-\beta}^{\beta}\int_{-\gamma}^{\gamma} u(x,y,z,t)\,u(x+\xi,y+\eta,z+\zeta,t)\,dx\,dy\,dz. \qquad (4\cdot1) \end{aligned}$$

Substituting from (3·3) for the two velocities in the integrand gives a multiple integral for R in terms of conditions at $t=0$. The integration connected with the x-direction is

$$\begin{aligned} &\frac{1}{2\alpha}\int_{-\alpha}^{\alpha} u(x,t)\,u(x+\xi,t)\,dx \\ &\quad = \frac{1}{2\alpha}\int_{-\alpha}^{\alpha}\int_{-\alpha}^{\alpha}\int_{-\alpha}^{\alpha} \frac{u(X,0)\,u(X',0)}{4\pi\nu t}\exp\left[\frac{-(X-x)^2-(X'-x-\xi)^2}{4\nu t}\right]dX\,dX'\,dx \\ &\quad \approx \frac{1}{2\alpha}\int_{-\alpha}^{\alpha}\int_{-\alpha}^{\alpha} \frac{u(X,0)\,u(X',0)}{2\sqrt{(2\pi\nu t)}}\exp\left[-\frac{(X-X'+\xi)^2}{8\nu t}\right]dX\,dX' \end{aligned}$$

if α is sufficiently large. Putting $X' = X+a$,

$$\begin{aligned} &= \frac{1}{2\alpha}\int_{-\alpha}^{\alpha}\int_{-\alpha-X}^{\alpha-X} \frac{u(X,0)\,u(X+a,0)}{2\sqrt{(2\pi\nu t)}}\exp\left[-\frac{(\xi-a)^2}{8\nu t}\right]dX\,da' \\ &= \frac{1}{2\sqrt{(2\pi\nu t)}}\int_{-\alpha-X}^{\alpha-X} \overline{u(X,0)\,u(X+a,0)}\exp\left[-\frac{(\xi-a)^2}{8\nu t}\right]da. \qquad (4\cdot2) \end{aligned}$$

The other two sets of integrations give a similar result, so that

$$\begin{aligned} R(\xi,\eta,\zeta,t) &= \frac{1}{(8\pi\nu t)^{\frac{3}{2}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(a,b,c,0) \\ &\qquad \times\exp\left[-\frac{(\xi-a)^2-(\eta-b)^2-(\zeta-c)^2}{8\nu t}\right]da\,db\,dc, \qquad (4\cdot3) \end{aligned}$$

where the limits of integration are again written as infinite without causing a significant error. Equation (4·3) is typical of the relations between mean values at two different instants in the final period of decay.

Our interest lies chiefly in the possibility that the mean values in the final period tend to asymptotic forms which are independent of the initial conditions. With this in mind, consider the expression (4·3) for large values of t. More precisely, let νt be large compared with values of $a^2+b^2+c^2$ ($= s^2$ say) for which $R(a,b,c,0)$ is large enough to contribute to the integral. The exponential will then be approximately unity, and the right side will be independent of ξ, η and ζ, unless $\xi^2+\eta^2+\zeta^2$ ($= r^2$ say) is of order νt. That is, the correlation coefficient at time t is approximately unity unless $r^2 = 0(\nu t)$. Hence $2(a\xi + b\eta + c\zeta)/8\nu t$ is of order $(\nu t)^{-\frac{1}{2}}$, $s^2/\nu t$ is of order $(\nu t)^{-1}$, and (4·3) can be written

$$R(\xi, \eta, \zeta, t) = \frac{e^{-r^2/8\nu t}}{(8\pi\nu t)^{\frac{3}{2}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(a, b, c, 0) \sum_n \frac{1}{n!}\left[\frac{2(a\xi + b\eta + c\zeta) - s^2}{8\nu t}\right]^n da\, db\, dc, \tag{4·4}$$

only the first few terms of the series being required.

There are two simple lemmas which will be used in evaluating the contributions from different terms in the series expansion. The first is that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(a, b, c, 0)\, db\, dc = 0, \tag{4·5}$$

since continuity requires the total flow across an infinite plane to be zero.* The second is that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a^i b^j c^k R(a, b, c, 0)\, da\, db\, dc = 0, \tag{4·6}$$

when $i + j + k$ is an odd integer. This follows from the requirement of homogeneity that $R(a, b, c, 0) = R(-a, -b, -c, 0)$.

It follows immediately from (4·5) that the leading term, $n = 0$, makes no contribution to the integral (4·4). The next largest term in the series, of order $(\nu t)^{-\frac{1}{2}}$, is obtained from $n = 1$, but (4·6) shows that it too makes no contribution to the integral. The first non-zero contribution to the integral comes from terms of order $(\nu t)^{-1}$ in the series, and hence, to the first approximation,

$$R(\xi, \eta, \zeta, t) = \frac{\pi e^{-r^2/8\nu t}}{(8\pi\nu t)^{\frac{3}{2}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(a, b, c, 0)\left[\frac{(a\xi + b\eta + c\zeta)^2}{4\nu t} - s^2\right] da\, db\, dc. \tag{4·7}$$

Equation (4·7) describes the asymptotic behaviour of the most general space-interval double-velocity correlation and shows that the initial conditions are relevant. But there are certain special cases, which also happen to be important, in which the effect of the initial conditions is confined, at most, to a multiplicative constant. The most important special case is $r = 0$, giving the asymptotic law of energy decay as

$$u'^2(t) = \frac{-\pi\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s^2 R(a, b, c, 0)\, da\, db\, dc}{(8\pi\nu t)^{\frac{3}{2}}}, \tag{4·8}$$

---

* More rigorously, note that if u', v' and w' are the velocity components at the point $(x + a, y + b, z + c)$, continuity requires

$$\overline{u(x, y, z)\left[\frac{\partial u'}{\partial a} + \frac{\partial v'}{\partial b} + \frac{\partial w'}{\partial c}\right]} = 0,$$

and hence $$\frac{\partial}{\partial a}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(a, b, c)\, db\, dc = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{\partial \overline{uv'}}{\partial b} + \frac{\partial \overline{uw'}}{\partial c}\right) db\, dc$$
$$= 0.$$

Thus the integral on the left side is constant, and on taking the particular case in which a is very large, the relation (4·5) follows.

in agreement with the deductions described in § 2. When $\eta = \zeta = 0$, $R(\xi, \eta, \zeta, t)$ degenerates to the longitudinal correlation previously denoted by $\overline{u^2}(t) f(\xi, t)$; hence, making use of (4·5) again,

$$\overline{u^2}(t) f(\xi, t) = -\frac{\pi e^{-\xi^2/8\nu t}}{(8\pi\nu t)^{\frac{5}{2}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s^2 R(a, b, c, 0)\, da\, db\, dc, \tag{4·9}$$

so that

$$f(\xi, t) = e^{-\xi^2/8\nu t}. \tag{4·10}$$

With the proviso that the integral in (4·8) and (4·9) is neither zero nor divergent, all longitudinal correlations should tend to the form (4·10) in the final period of decay.

It should be noticed that no assumptions about symmetry of the turbulence have yet been made, and the result (4·10) in no way depends on isotropy. The only assumption made is that the turbulence is approximately homogeneous. In point of fact, in a field of uniform mean flow a state of isotropy will usually have been attained before the decay is sufficiently far advanced for the laws (4·8) and (4·10) to apply, so that this generality may not have much practical use. When the turbulence at the initial instant $t = 0$ is isotropic, the triple integral in (4·8) and (4·9) can be reduced:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s^2 R(a, b, c, 0)\, da\, db\, dc \\
&\qquad = \overline{u^2}(0) \int_0^{\infty} \int_0^{\pi} \int_0^{2\pi} s^2 \left[f(s, 0) + \frac{s}{2} \frac{\partial f(s, 0)}{\partial s} \sin^2\theta \right] s^2 \sin\theta\, ds\, d\theta\, d\phi \\
&\qquad = 4\pi \overline{u^2}(0) \int_0^{\infty} s^4 \left[f(s, 0) + \frac{s}{3} \frac{\partial f(s, 0)}{\partial s} \right] ds \\
&\qquad = -\frac{8\pi}{3} \overline{u^2}(0) \int_0^{\infty} s^4 f(s, 0)\, ds \\
&\qquad = -\frac{8\pi}{3} \Lambda,
\end{aligned}$$

so that (4·8) becomes

$$\overline{u^2}(t) = \frac{\Lambda}{48\sqrt{(2\pi)}} (\nu t)^{-\frac{5}{2}}. \tag{4·11}$$

It is also possible to deduce the asymptotic form of the time-interval double-velocity correlation, i.e. the correlation between parallel velocity components at the same point but at two different times separated by an interval τ. A suitable definition is

$$\begin{aligned}
R(\tau, t) &= \overline{u(x, y, z, t - \tfrac{1}{2}\tau)\, u(x, y, z, t + \tfrac{1}{2}\tau)} \\
&= \frac{1}{8\alpha\beta\gamma} \int_{-\alpha}^{\alpha} \int_{-\beta}^{\beta} \int_{-\gamma}^{\gamma} u(x, y, z, t - \tfrac{1}{2}\tau)\, u(x, y, z, t + \tfrac{1}{2}\tau)\, dx\, dy\, dz.
\end{aligned} \tag{4·12}$$

Substituting from (3·3), the integration connected with the x-direction is

$$\begin{aligned}
&\frac{1}{2\alpha} \int_{-\alpha}^{\alpha} \int_{-\alpha}^{\alpha} \int_{-\alpha}^{\alpha} \frac{u(X, 0)\, u(X', 0)}{4\pi\nu\sqrt{(t^2 - \frac{1}{4}\tau^2)}} \exp\left[-\frac{(X - x)^2}{4\nu(t - \frac{1}{2}\tau)} - \frac{(X' - x)^2}{4\nu(t + \frac{1}{2}\tau)} \right] dx\, dX\, dX' \\
&\qquad \approx \frac{1}{2\alpha} \int_{-\alpha}^{\alpha} \int_{-\alpha}^{\alpha} \frac{u(X, 0)\, u(X', 0)}{2\sqrt{(2\pi\nu t)}} \exp\left[-\frac{(X - X')^2}{8\nu t} \right] dX\, dX',
\end{aligned}$$

and τ has disappeared from the analysis. Hence, regarding the reference to the point x, y, z as understood,

$$R(\tau, t) = \overline{u(t-\tfrac{1}{2}\tau)\,u(t+\tfrac{1}{2}\tau)} = \overline{u^2(t)}. \qquad (4\cdot13)$$

When t is sufficiently large, (4·11) is valid so that the asymptotic form of the time-interval correlation coefficient is

$$\frac{\overline{u(t-\frac{1}{2}\tau)\,u(t+\frac{1}{2}\tau)}}{[\overline{u^2(t-\frac{1}{2}\tau)}\,\overline{u^2(t+\frac{1}{2}\tau)}]^{\frac{1}{2}}} = \frac{(t^2-\frac{1}{4}\tau^2)^{\frac{5}{4}}}{t^{\frac{5}{2}}} = \left(1-\frac{\tau^2}{4t^2}\right)^{\frac{5}{4}}. \qquad (4\cdot14)$$

Although (4·13) is valid for all values of τ such that $t-\frac{1}{2}\tau > 0$, (4·14) requires $[\nu(t-\frac{1}{2}\tau)]^{\frac{1}{2}}$ to be large compared with the spread of the spatial correlation at $t = 0$. In a practical experiment τ/t would normally be small enough for (4·14) to be written as $1-\frac{5}{16}(\tau^2/t^2)$.

The form of this time interval correlation throws a little more light on the nature of the final period. Equation (4·13) shows that, roughly speaking, the only change in the velocity at any point is a gradual diminution corresponding to the decay of the whole pattern of turbulence. The shape of the spatial distribution of velocity remains approximately the same as t increases, the steeper gradients being modified most. The turbulence may be described as aperiodic (in time), like other flow fields dominated by viscosity.

Finally, the most general double velocity correlation coefficient involving both space and time intervals may be found, by analysis similar to that leading to (4·13), to be a simple product of the correlation coefficients already discussed. Since

$$\overline{u(x,y,z,t-\tfrac{1}{2}\tau)\,u(x+\xi,y+\eta,z+\zeta,t+\tfrac{1}{2}\tau)} = \overline{u(x,y,z,t)\,u(x+\xi,y+\eta,z+\zeta,t)}$$

it follows that

$$\frac{\overline{u(x,y,z,t-\frac{1}{2}\tau)\,u(x+\xi,y+\eta,z+\zeta,t+\frac{1}{2}\tau)}}{[\overline{u^2(t-\frac{1}{2}\tau)}\,\overline{u^2(t+\frac{1}{2}\tau)}]^{\frac{1}{2}}} = \frac{R(\tau,t)}{[\overline{u^2(t-\frac{1}{2}\tau)}\,\overline{u^2(t+\frac{1}{2}\tau)}]^{\frac{1}{2}}}\,\frac{R(\xi,\eta,\zeta,t)}{\overline{u^2(t)}}, \qquad (4\cdot15)$$

and the asymptotic form may be obtained from the results already derived. In particular, the longitudinal correlation coefficient tends to the form

$$\left(1-\frac{\tau^2}{4t^2}\right)^{\frac{5}{4}} e^{-\xi^2/8\nu t}. \qquad (4\cdot16)$$

5. MEASUREMENTS OF TURBULENCE IN THE FINAL PERIOD

To produce isotropic turbulence whose decay may be observed over a sufficient interval for the establishment of the final period of decay, it is necessary to use a grid of very small mesh. It is known that the initial period of decay extends to about 150-mesh lengths from the grid, and a decay range much greater than this is required if a reasonable range of the final period is to be available after the transitional period. A mesh of 0·159 cm. was used in the experiments, allowing measurements to 1040-mesh lengths downstream from the grid. This grid is of the biplane type, and made of stretched eureka wires of diameter 0·030 cm. It is geometrically similar to grids previously used (Batchelor & Townsend 1948), with a spacing-diameter ratio of 5·33. Measurements were made of the turbulent intensity, of the dissipation parameter λ, and of the longitudinal double velocity correlation at an air speed of 615 cm.sec.$^{-1}$, and of the turbulent intensity at air speeds of 895 and 1280 cm.sec.$^{-1}$. The respective mesh Reynolds numbers, $R_M = UM/\nu$, are 650, 950 and 1360. Using a 0·635 cm. grid at 150 cm.sec.$^{-1}$ ($R_M = 635$), rather more accurate measurements in the initial period are possible and these are included in the results.

The technique for the measurement of the turbulent intensity and the length λ has been described previously (Townsend 1947), the only significant change here being the upward restriction of the amplifier frequency response in order that an adequate signal-noise ratio should be available when measuring the differentiated signal.

The scale of the turbulence produced by a 0·159 cm. mesh is so small that the distance between two hot-wires with finite velocity correlation is difficult to measure accurately. It was therefore decided to measure the auto-correlation function of the output of a single hot-wire, and to compute from it the longitudinal double-velocity correlation. If $F(\tau)$ is the auto-correlation for a time interval τ, then the longitudinal correlation is related to it by

$$f(r) = F(U\tau),$$

provided that the turbulence does not change appreciably (apart from its translation with velocity U) in a time r/U. This assumption is nearly true if the turbulence level is low, as has been shown by the experimental verification of the relation between the spectrum and the correlation function of isotropic turbulence (Taylor 1938). It is even more likely to be accurate under the conditions of these experiments, viz. very low turbulent intensities and weak diffusive properties.

To measure the auto-correlation, the hot-wire anemometer output must be compared at times separated by a constant but adjustable interval, which is done by delaying the transmission of the fluctuations. The anemometer output is recorded at two points on a magnetic tape driven at constant speed, and each recorded signal is reproduced after a time interval determined by the tape speed and the distance between each recording point and the corresponding reproducing point as measured along the tape. The correlation between these reproduced signals is measured, and the appropriate correlation interval is the difference of the time delays of the two reproduced signals. In this way, not only are effects of slight distortion minimized, but both positive and negative time delays can be used. A detailed description of the magnetic tape time delay equipment will be published separately.

A survey of the turbulence behind the 0·159 cm. mesh showed that, although the mesh appeared completely uniform to the eye, considerable spatial variations of intensity occurred across the tunnel, and it became evident that these were caused by slight non-uniformity of the mesh. The authors' notion is that at Reynolds numbers close to values which are critical for the instability of the laminar flow immediately behind the grid, small large-scale irregularities in the grid will cause the production of widely different levels of turbulence behind different parts of the grid. That the effect is due to large-scale irregularities is confirmed by the comparative insensitivity of the intensity distribution to quite large small-scale irregularities, such as that produced by omitting one element of the mesh. This intensity variation, although it is annoying, is not a serious obstacle, for the turbulence shows extremely weak diffusive properties, and the transverse intensity profile remains nearly similar during a decay of intensity in a ratio twenty to one. This is evidence that the turbulence consists of eddies nearly independent of one another. The effect of non-uniformity has been removed from the results by averaging the intensity over a considerable part of the tunnel section. It is probable that the mean intensity obtained in this way is a good approximation to the intensity behind an ideal mesh.

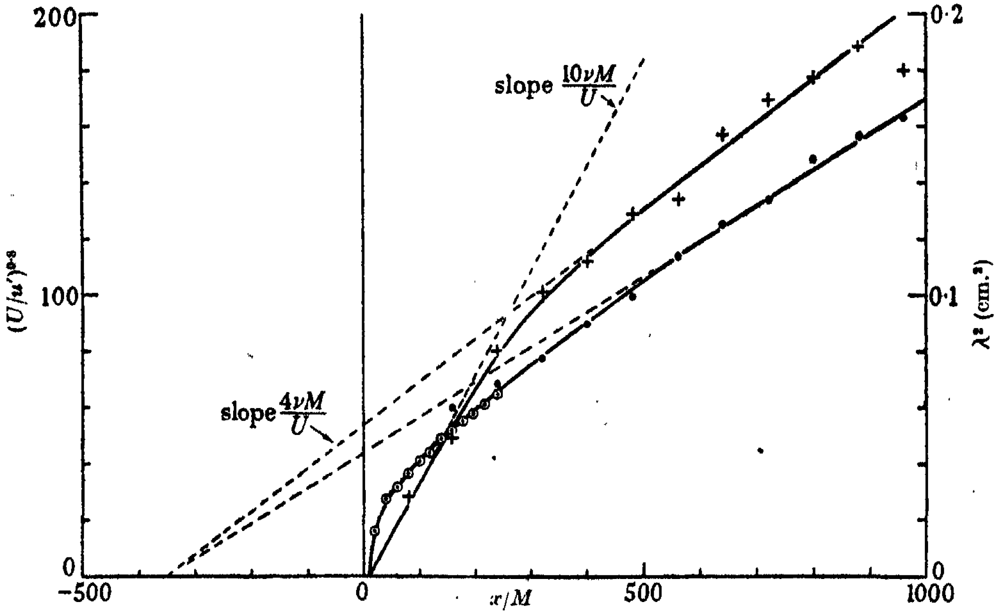


FIGURE 1. Variation of intensity and λ ($R_M = 650$).

● $(U/u')^{0\cdot8}$: $M = 0\cdot159$ cm., $U = 620$ cm.sec.$^{-1}$.
⊙ $(U/u')^{0\cdot8}$: $M = 0\cdot635$ cm., $U = 150$ cm.sec.$^{-1}$.
+ λ^2: $M = 0\cdot159$ cm., $U = 620$ cm.sec.$^{-1}$.

In figure 1, $(U^2/\overline{u^2})^{0\cdot4}$ and λ^2 measured at a Reynolds number of 650 are plotted as functions of x/M. The foregoing theory predicts the following asymptotic relations in the final period:

$$\left(\frac{U^2}{\overline{u^2}}\right)^{0\cdot4} = A\,\frac{(x-x_0)}{M}, \quad \lambda^2 = \frac{4\nu}{U}(x-x_0),$$

where A and x_0 are constants characteristic of the flow, and

$$A = \frac{4\nu}{MU}\left(\frac{U^2M^5}{\overline{u^2}\lambda^5}\right)^{\frac{2}{5}} = \frac{\nu}{MU}\left(\frac{48\sqrt{(2\pi)}\,U^2M^5}{\Lambda}\right)^{\frac{2}{5}}.$$

On the diagram, lines satisfying these equations are drawn to fit the experimental points, and good agreement for $x/M > 400$ is obtained if $A = 0\cdot129$ and $x_0 = -350M$.

For reference, figure 1 shows a line representing the known variation of λ^2 in the initial period (Batchelor & Townsend 1948), viz.

$$\lambda^2 = \frac{10\nu}{U}(x - 10M),$$

and this adequately represents the range to $x/M = 160$. At this Reynolds number then, the asymptotic energy decay relation in the final period is established after 400-mesh lengths.

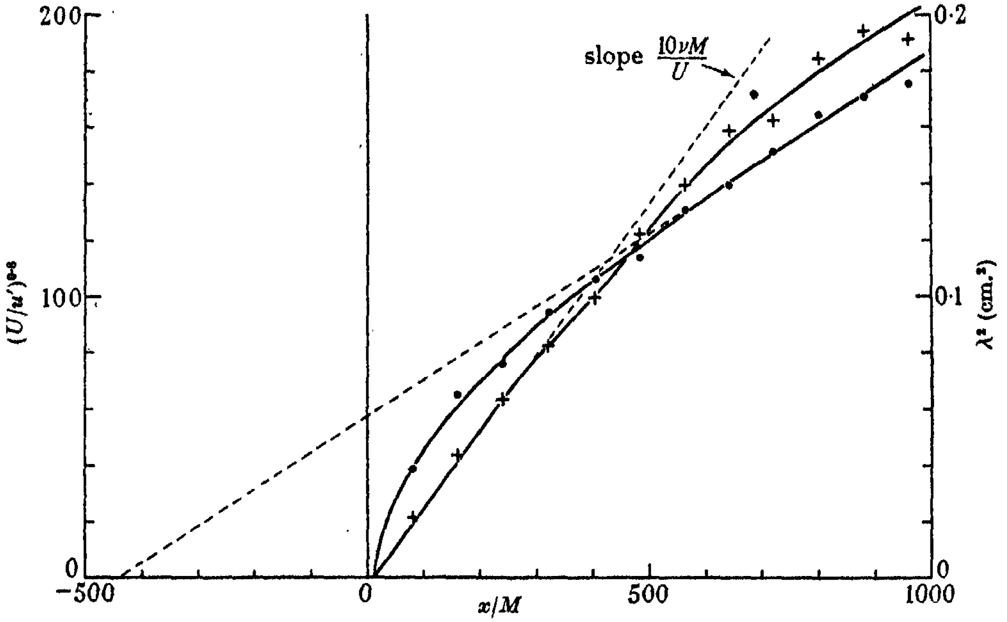


FIGURE 2. Variation of intensity and λ ($R_M = 950$). • $(U/u')^{0.8}$; + λ^2.

Measurements of intensity and of λ^2 at higher speeds are given in figures 2 and 3, but experimental difficulties in the measurements of λ at high speeds and low intensities make the results for λ^2 at large x/M rather irregular. It is evident from the intensity plots that the setting up of the final period asymptotic decay cannot occur here before $x/M = 600$ and is probably not complete before $x/M = 800$. It can be said that the slope of the λ^2 versus x/M curve is less than $10\nu(M/U)$, except in the initial period, but the interesting values for $x/M > 700$ are subject to considerable uncertainty. Supposing the asymptotic decay law toward the limits of observation, we find the following values for A, x_0/M and Λ.*

mesh Reynolds number	A	x_0/M	Λ (cm.^{7} sec.$^{-2}$)
650	0·13	−350	0·072
950	0·13	−440	0·058
1360	0·13	−450	0·048

---

* The results for Λ show that at a given value of x/M lying in the initial period, the value of $\int_0^\infty r^4 f(r)\,dr$ diminishes as U increases; if A were accurately constant, this decrease would be as $U^{-\frac{1}{2}}$, which is of interest in connexion with the question of the turbulence set up immediately behind a grid.

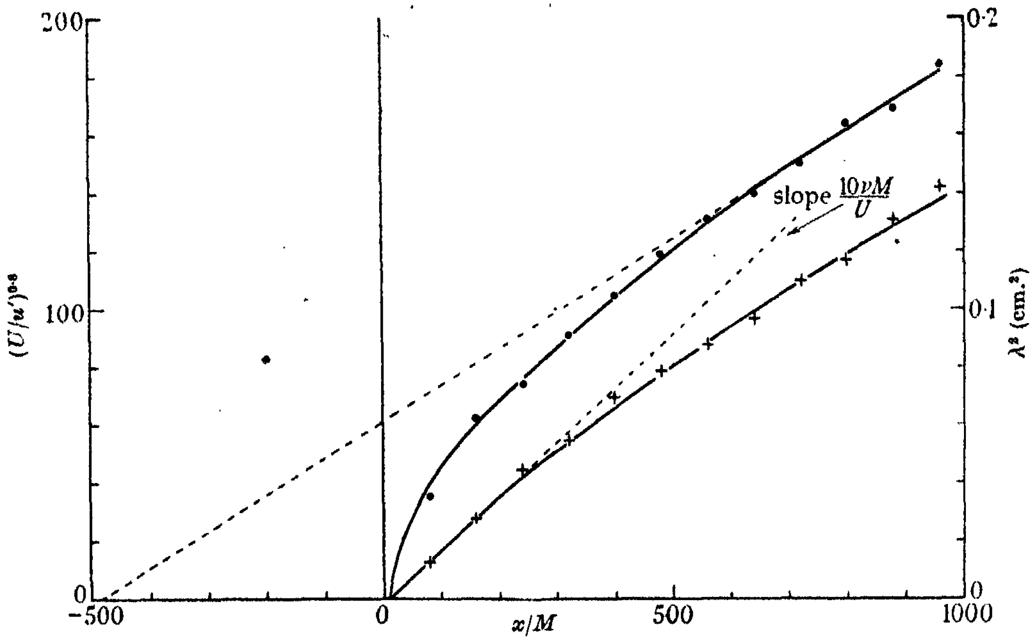


FIGURE 3. Variation of intensity and λ ($R_M = 1360$). • $(U/u')^{0.8}$; + λ^2.

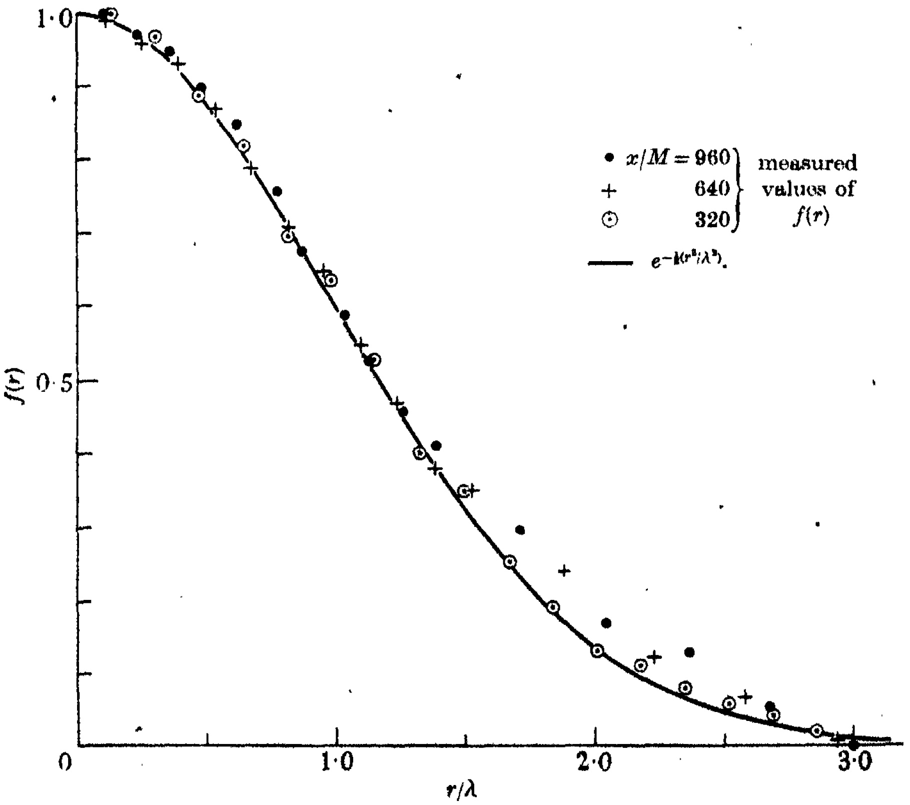


FIGURE 4. Double correlation function ($M = 0{\cdot}159$ cm., $U = 620$ cm.sec.$^{-1}$, $R_M = 650$).

The double correlation function is plotted in figure 4 as a function of r/λ for several decay times, the value of λ being taken from the radius of curvature at $r = 0$ in each case. Also shown is the theoretical asymptotic form of the function, viz.

$$f(r) = e^{-\frac{1}{2}r^2/\lambda^2}.$$

The agreement is good and adequately confirms the existence of the predicted correlation function.

6. Nature of final period decay and conditions for its existence

The present experiments confirm the existence of final period decay of isotropic turbulence for $x/M > 400$ at a mesh Reynolds number of 650, and verify the theoretical predictions of decay and correlation. The final period flow may be described as an aperiodic approach to complete rest. Asymptotic forms of the decay law and of the correlation functions may be inferred, even though neither the details nor the statistical characteristics of the initial motion are known. It has been found that, during the final period, turbulent energy diffuses very slowly down an intensity gradient, and that the time correlation is close to unity over a considerable range of time. This all points to a high stability and uniformity of structure of the motion, and it is likely that the turbulence consists of almost independent eddies. Such eddies would lose energy by viscous dissipation, and simultaneously would increase in size by viscous spreading, occasionally amalgamating with an adjacent eddy but only rarely. The limited extent of the spatial correlation suggests that the shape and structure of the eddies is simple, and that they have a limited range of sizes. Confirmation of these surmises would be very difficult to obtain, as the only practical method would seem to be the measurement of three- or four-point correlations.

In order that final period decay should be possible, it is necessary that inertia forces should play a negligible part in the turbulent motion. For the asymptotic relations to apply it is then necessary that decay shall have progressed so far that the initial conditions have been obliterated by viscous action. To express the first condition exactly requires a knowledge of the turbulent structure, but, if it can be assumed that $S_0, = \overline{\left(\frac{\partial u}{\partial x}\right)^3} \Big/ \left[\overline{\left(\frac{\partial u}{\partial x}\right)^2}\right]^{\frac{3}{2}}$, has the same value at *all* Reynolds numbers (Batchelor & Townsend 1947), then the condition that inertia plays no part in determining the vorticity balance is that

$$R_\lambda \ll -\frac{30}{7}\frac{1+n}{nS_0}$$
$$= 15$$

if the decay law is such that $\overline{u^2} \propto t^{-\frac{5}{2}}$, i.e. $n = \frac{5}{2}$. If the correlation function is of normal form, the second condition will include the restriction

$$4\nu t \gg L_0^2, \quad \text{i.e.} \quad \frac{x}{M} \gg \frac{R_M}{4}\left(\frac{L_0}{M}\right)^2,$$

where L_0 is the scale of the turbulence, possibly the integral scale $\int_0^\infty f(r)\,dr$, during the initial period. Taking $L_0/M = 0{\cdot}95$ as representative of conditions at $x/M = 100$, the condition becomes $x/M \gg 0{\cdot}23R_M$. At $R_M = 650$, this lower bound is 150, whereas observation shows the asymptotic relations to occur at $x/M = 400$. Both the above conditions are necessary rather than sufficient.

By considering the behaviour of R_λ during decay, it is possible to make some deductions about the length of the transitional period. If the decay is represented as proceeding in the initial period (with constant R_λ) for a time t_0, and then in the transitional period with a decay law intermediate between that in the initial and final periods, say

$$\lambda^2 = 7\nu(t+\tfrac{3}{7}t_0), \quad \overline{u^2} \propto (t+\tfrac{3}{7}t_0)^{-\frac{10}{7}},$$

the value at any time in the transition period ($t > t_0$) will be

$$R_\lambda = (R_\lambda)_{\text{initial period}}\left(\frac{7}{10}\frac{t}{t_0}+\frac{3}{10}\right)^{-\frac{3}{14}}.$$

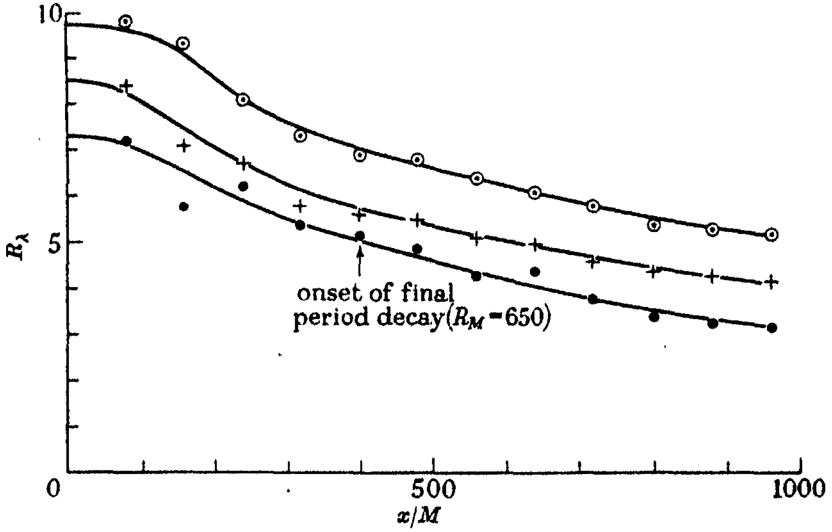


FIGURE 5. Variation of R_λ with x/M. Values of R_M: • 650, + 950, ⊙ 1360.

This means that the decrease of R_λ with time is extremely slow, and if $(R_\lambda)_{\text{initial period}}$ is much greater than 15, it is most improbable that the final period can be attained in a practicable decay time, e.g. a reduction of R_λ by a factor of 4 would require decay for a time of order $1000t_0$. Since the initial period usually extends for about 150 mesh lengths, the reason why previous wind tunnel experiments have not detected final period decay relations is apparent.

In figure 5, values of R_λ calculated from the results in figures 1 to 3 are plotted against x/M. The expected behaviour is an initial range where R_λ is constant followed by a more or less gradual decrease. As nearly as can be made out, at the lowest Reynolds number in the experiments ($R_M = 650$), the final period decay law sets in at about $x/M = 400$ where $R_\lambda \approx 5$. At the next higher Reynolds number, $R_M = 950$, final period decay was not certainly observed, although the value, $R_\lambda = 5$, is attained when $x/M = 650$. Although the use of a simple Reynolds number criterion for the existence of final period flow is in general unjustifiable, there is nothing in the present results against the empirical, necessary criterion, $R_\lambda < 5$.

The experiments described in this paper were carried out in the Cavendish Laboratory. The wind tunnel and hot-wire equipment were acquired with the aid of a grant from the Aeronautical Research Council. During the course of the work one of us (G.K.B.) held a Senior Studentship given by the Royal Commissioners for the Exhibition of 1851.

REFERENCES

Batchelor, G. K. 1948 *Quart. Applied Math.* **6**, 97–116.
Batchelor, G. K. & Townsend, A. A. 1947 *Proc. Roy. Soc.* A, **190**, 534–550.
Batchelor, G. K. & Townsend, A. A. 1948 *Proc. Roy. Soc.* A, **193**, 539.
Kármán, T. von & Howarth, L. 1938 *Proc. Roy. Soc.* A, **164**, 192–215.
Loitsiansky, L. 1939 *Cent. Aero. and Hydro. Inst.* (*Moscow*), *Rep.* no. 440. (Translated as *N.A.C.A. Tech. Memo.* no. 1079.)
Millionshtchikov, M. 1939 *C.R. Acad. Sci. U.R.S.S.* **22**, 231–237.
Reissner, E. 1938 *Proc. 5th Int. Congr. Appl. Mech.* pp. 359–361.
Sedov, L. 1944 *C.R. Acad. Sci. U.R.S.S.* **42**, 116–119.
Taylor, G. I. 1938 *Proc. Roy. Soc.* A, **164**, 476.
Townsend, A. A. 1947 *Proc. Camb. Phil. Soc.* **43**, 560.

On the classical theory of particles

BY C. JAYARATNAM ELIEZER, *Christ's College, University of Cambridge*

(*Communicated by R. E. Peierls, F.R.S.—Received 7 February* 1948)

A set of classical relativistic equations of motion of an electron in an electromagnetic field is postulated. These equations are free from 'run-away' solutions, and give the same results as the Maxwell-Lorentz theory for non-relativistic motions when the external electromagnetic field does not vary too rapidly. For the scattering of light by an electron, the scattering cross-section is independent of the frequency and is a universal constant. This brings out a point of difference from the Lorentz-Dirac equations according to which the scattering cross-section varies inversely as the square of the frequency of the incident light, for large frequencies. For the motion of an electron towards a fixed proton, the equations allow a collision, unlike the Lorentz-Dirac equations according to which the electron is brought to rest before it reaches the proton.

1. INTRODUCTION

There has been in recent years a marked revival of interest in the classical theories of particles. After the pioneering work of Abraham and Lorentz in this field some forty or fifty years ago, comparatively little attention was paid until very recently to the development of the classical theory, the centre of research in theoretical physics being occupied by the rapidly expanding field of the quantum theory. However, the various difficulties that have now arisen in quantum electrodynamics, particularly that of infinite self-energies of point particles, have led to the view that some of these difficulties may be of classical origin, and that a promising method of approach to eliminate these difficulties would be to study and improve the classical theory first, before passing on to the quantum theory.

A substantial step in this direction was made by Dirac (1938) when he showed that the infinite self-energy, ascribed to the point electron by the Maxwell-Lorentz classical theory of electromagnetism, may be subtracted out in a Lorentz invariant way. The subtraction process led to a well-defined scheme of equations of motion of an electron in an electromagnetic field. These equations are consistent with the conservation laws and with the principle of special relativity. These equations in their non-relativistic approximation were found to be the same as the equations obtained earlier by Lorentz (1909). Lorentz regarded the electron as a small charged sphere of radius r_0 and obtained for the force on the electron due to its own field an expression in the form of a series in ascending powers of r_0. The first term of the series is $e^2\dot{\mathbf{v}}/r_0$, where e is the electronic charge and $\mathbf{v}$ its velocity; dots denote differentiation with respect to the time, and units are chosen so that the velocity of light is unity. This first term describes electromagnetic inertia and may be added on to any mechanical inertia the electron may have, the two together being then represented by $m\dot{\mathbf{v}}$ in the equations of motion, where m is the observed mass of the electron. The second term of the series is $\frac{2}{3}e^2\ddot{\mathbf{v}}$ and is independent of the shape of the electron. It accounts for the loss of energy by radiation at a rate $\frac{2}{3}e^2\dot{\mathbf{v}}^2$. The higher terms of the series depend on the structure of the electron, and being successively proportional to $r_0, r_0^2, \ldots$ are regarded as small since r_0 is small. The Lorentz equations of motion for an electron of mass m and charge e in an electromagnetic field described by the field vectors $\mathbf{E}$ and $\mathbf{H}$ are thus

$$m\dot{\mathbf{v}} - \tfrac{2}{3}e^2\ddot{\mathbf{v}} = e(\mathbf{E} + \mathbf{v} \times \mathbf{H}). \tag{1}$$

These equations were rederived by Dirac by using a point model for the electron and using the subtraction process; but whereas Lorentz's method of derivation makes them necessarily approximate, Dirac has suggested that according to his derivation there are hopes that these equations may be exact within the limits of the classical theory.

2. SELF-ACCELERATING MOTIONS. POSSIBLE WAYS OF ELIMINATING THEM

Although the difficulty of infinite self-energy is satisfactorily accounted for by Dirac's theory there arise, however, other difficulties. The equations of motion are found to have solutions which do not correspond to motions that are observed physically. Equation (1) involves $\ddot{\mathbf{v}}$ and the solution involves more arbitrary constants than are necessary to fix the actual motion. Thus in the absence of any incident field the motion is given by

$$m\dot{\mathbf{v}} - \tfrac{2}{3}e^2\ddot{\mathbf{v}} = 0, \tag{2}$$

and hence by

$$\mathbf{v} = \mathbf{A} + \mathbf{B}e^{at}, \tag{3}$$

where $a = 3m/2e^2$, and $\mathbf{A}$ and $\mathbf{B}$ are arbitrary constant vectors. If $\mathbf{B}$ is not zero the electron would rapidly build up a velocity and radiate energy at a rapid rate. Such a motion has not been observed. Similar non-physical solutions exist for motion in other electromagnetic fields.

There seem to be three possible ways of attempting to meet this difficulty:

(i) One is to suppose that out of the entire family of solutions of the equations of motion one should pick out, by the use of some extra physical condition, only those which are physically allowable. Thus for the above case of the free electron the extra condition is imposed that the acceleration should be zero, and the allowed solution corresponds to uniform motion in agreement with experiment (Dirac 1938). This condition may be suitably extended to homogeneous electric or magnetic fields (Eliezer & Mailvaganam 1945).

But there are difficulties in this method of excluding the non-physical solutions. There are some problems for which there do not seem to exist adequate physical solutions at all. For the electron in the hydrogen atom (Eliezer 1943), or an electron in the field of a thin charged plate (Eliezer 1945), all the solutions of the equations appear to have non-physical characteristics. This may not be a serious difficulty since adequate physical solutions may exist in the corresponding quantum theory, but there would then be the difficulty that the procedure of selecting some particular solution out of a family of solutions does not fit in with the general principles of quantum mechanics as at present formulated (Dirac 1946). It may not be possible to exclude quantum transitions between the physical and the non-physical motions.

(ii) Another way of eliminating this difficulty of non-physical solutions is to derive equations of motion which do not contain higher derivatives of velocity than the first derivative $\dot{\mathbf{v}}$. This is possible if the field of a moving electron is taken to be the mean of its retarded and advanced fields (Eliezer 1947). This assumption leads to the usual equations of motion without the radiation damping terms, but according to this derivation these equations are consistent with the conservation laws. A possible objection to this theory is that in leading to equations which do not contain the damping terms it does not account for the experimentally observed radiation loss of an electron. This would be a serious objection if it can be conclusively shown that the energy loss is a classical effect and that it holds for a single electron. It may be that only the many-body problem is encountered in nature and that the observed radiation damping has its origin in the complicated way in which the electrons are acted upon by the retarded and advanced fields of the other electrons. Wheeler & Feynman (1945) have attempted to build up a theory of radiation in which the force of radiation reaction on a particle is due to the field of the particles of the absorbing medium surrounding it; but it does not seem clear whether this attempt has been successful. Again, even if the observed effect applies to a single electron the radiation loss may perhaps be accountable by a quantum theory which takes over classical equations which do not contain the damping terms. Evidence on these questions seem to me to be indecisive.

(iii) A third possible approach to eliminate the difficulty of self-acceleration is to keep the extended model of the electron. The higher terms of the series for the force of radiation reaction should then be taken into account. In all applications of the Lorentz equations these higher terms have been generally assumed to be small. But for the self-accelerating motions discussed above these higher terms are not small. According to the solution (3), $\mathbf{v}$, $\dot{\mathbf{v}}$, $\ddot{\mathbf{v}}$, ... are seen to increase successively by a factor a so that $\mathbf{v}$, $\dot{\mathbf{v}}/a$, $\ddot{\mathbf{v}}/a^2$, ... are all of the same order ($1/a$ being the order of the radius of the electron).

In this paper I intend to study certain equations of motion with the higher terms included. These higher terms depend on the distribution of charge of the electron, and a particular distribution which would lead to simple equations of motion will be taken. These equations are expressed in relativistic form and are taken as the starting point for a classical theory of electrons.

3. THE EQUATIONS OF MOTION

We consider the motion of an electron in an electromagnetic field. The Lorentz-Dirac equations give rise to self-accelerating motions whether we use the relativistic form of the equations or their non-relativistic approximation. Therefore in trying to eliminate these difficulties it seems the right procedure to start with the simpler non-relativistic case. We shall suppose that the electron starts with an initial motion in which the velocity, acceleration and the higher derivatives of velocity are small so that products of two or more of them may be neglected. The crucial point is to have a system of equations such that for a particle moving under no forces and with an initial motion in which the velocity and its derivatives are small, the solution of the equations of motion does not require the particle to work itself up to a rapidly increasing velocity and so contradict the conditions under which the equations are assumed to hold. If such a scheme of equations could be found, then it would provide a hopeful starting point for a consistent theory.

Page (1918) has investigated the form of the higher order terms in the equations of motion. If the electron is taken to be a charged sphere of radius r_0 and total charge e, the charge being confined to the surface of the sphere, then the equations of motion of the electron in an electric field $\mathbf{E}$ are

$$m\mathbf{f} - \tfrac{2}{3}e^2\dot{\mathbf{f}} + \frac{2e^2}{3r_0}\sum_{n=2}^{\infty}(-1)^n \frac{(2r_0)^n \mathbf{f}^{(n)}}{(n+1)!} = e\mathbf{E}, \tag{4}$$

where $\mathbf{f}^{(n)}$ denotes the nth derivative of the acceleration $\mathbf{f}$ with respect to the time t, and m is the sum of the electron's mechanical mass (if any) and its electromagnetic mass $\frac{2}{3}e^2/r_0$. This equation has some interesting features, but they cause some complications and they will be discussed in a later paper. If we consider other distributions of charge, the equations of motion will then have the form

$$m\mathbf{f} - \tfrac{2}{3}e^2\dot{\mathbf{f}} + \sum_{n=2}^{\infty}(-1)^n c_n \mathbf{f}^{(n)} = e\mathbf{E}, \tag{5}$$

where the c's are positive constants which depend on the choice of the charge distribution. For a point electron the c's are zero.

We take (for reasons of simplicity which will become apparent later) a model which is such that

$$c_n = m/a^n \quad (n \geqslant 2), \tag{6}$$

where $a = 3m/2e^2$ as before. The radius of the electron is of the order $1/a$. Thus we take as the equations of motion

$$m\left[\mathbf{f} - \frac{1}{a}\mathbf{f}^{(1)} + \frac{1}{a^2}\mathbf{f}^{(2)} - \ldots + (-1)^n\frac{1}{a^n}\mathbf{f}^{(n)} + \ldots\right] = e\mathbf{E}, \tag{7}$$

that is,

$$\left[1 - \frac{D}{a} + \frac{D^2}{a^2} - \ldots + (-1)^n\frac{D^n}{a^n} + \ldots\right]m\mathbf{f} = e\mathbf{E}, \tag{8}$$

where $D = d/dt$. The nature of the model corresponding to these equations is not discussed here, as such a discussion would be rather complicated, and also because we shall be looking at these equations from a different point of view later. We have to assume that the choice (6) gives a consistent model.

The series on the left-hand side of (7) is convergent only when $\mathbf{f}^{(n)}$ tends to zero as n tends to infinity. It is then easy to show that the only solution of (7) satisfying this requirement is

$$m\mathbf{f} = \left(1+\frac{D}{a}\right)e\mathbf{E}. \tag{9}$$

Equation (7) and its solution (9) are valid only when the external field $\mathbf{E}$ is such that $(D/a)^n\,\mathbf{E}\to\mathbf{0}$ as $n\to\infty$.

4. The energy equation

The energy equation may be obtained by taking the scalar product of (7) with $\mathbf{v}$, thus obtaining

$$m\left[\mathbf{v}\,.\,\mathbf{f}-\frac{1}{a}\mathbf{v}\,.\,\dot{\mathbf{f}}+\ldots+(-1)^n\frac{1}{a^n}\mathbf{v}\,.\,\mathbf{f}^{(n)}+\ldots\right] = e\mathbf{E}\,.\,\mathbf{v}. \tag{10}$$

Now

$$\mathbf{v}\,.\,\mathbf{f}^{(2n-1)} = \frac{d}{dt}[\mathbf{v}\,.\,\mathbf{f}^{(2n-2)}-\mathbf{f}\,.\,\mathbf{f}^{(2n-3)}+\ldots+(-1)^{n-1}\mathbf{f}^{(n-2)}\,.\,\mathbf{f}^{(n-1)}]+(-1)^n\mathbf{f}^{(n-1)^2}, \tag{11}$$

$$\mathbf{v}\,.\,\mathbf{f}^{(2n)} = \frac{d}{dt}[\mathbf{v}\,.\,\mathbf{f}^{(2n-1)}-\mathbf{f}\,.\,\mathbf{f}^{(2n-2)}+\ldots+(-1)^{n-1}\mathbf{f}^{(n-2)}\,.\,\mathbf{f}^{(n)}+\tfrac{1}{2}(-1)^n\mathbf{f}^{(n-1)^2}]. \tag{12}$$

Therefore equation (10) may be written

$$\begin{aligned}\frac{d}{dt}\Bigg[V+\tfrac{1}{2}m\mathbf{v}^2+m\sum_{n=1}^{\infty}\Bigg\{&-\frac{1}{a^{2n-1}}\sum_{r=0}^{n-1}(-1)^r\,\mathbf{v}^{(r)}\,.\,\mathbf{v}^{(2n-r-1)}\\&+\frac{1}{a^{2n}}\left(\sum_{r=0}^{n-1}(-1)^r\,\mathbf{v}^{(r)}\,.\,\mathbf{v}^{(2n-r)}+\tfrac{1}{2}(-1)^n\,\mathbf{v}^{(n)^2}\right)\Bigg\}\Bigg] = -m\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{a^{2n-1}}\mathbf{v}^{(n)^2},\end{aligned} \tag{13}$$

where V is the potential energy of the electron in the field $\mathbf{E}$. Writing

$$\begin{aligned}W = V+\tfrac{1}{2}m\mathbf{v}^2+m\sum_{n=1}^{\infty}\Bigg\{&-\frac{1}{a^{2n-1}}\sum_{r=0}^{n-1}(-1)^r\,\mathbf{v}^{(r)}\,.\,\mathbf{v}^{(2n-r-1)}\\&+\frac{1}{a^{2n}}\left(\sum_{r=0}^{n-1}(-1)^r\,\mathbf{v}^{(r)}\,.\,\mathbf{v}^{(2n-r)}+\tfrac{1}{2}(-1)^n\,\mathbf{v}^{(n)^2}\right)\Bigg\},\end{aligned} \tag{14}$$

we may interpret W as the total energy of the electron. The first term gives the potential energy, the second term the usual kinetic energy, and the remaining terms correspond to energy due to the acceleration and its derivatives. If $\mathscr{R}$ denotes the rate of loss of energy, then

$$\mathscr{R} = -\frac{dW}{dt} = m\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{a^{2n-1}}\mathbf{v}^{(n)^2}. \tag{15}$$

The first term of the series is $m/a\dot{\mathbf{v}}^2 = \tfrac{2}{3}e^2\dot{\mathbf{v}}^2$, which is the usual expression for the rate of loss of energy of a point electron. The remaining terms arise because we are using a finite model for the electron.

5. SOME APPLICATIONS

(i) *Free electron*

For an electron in the absence of any incident field, the equation of motion (9) gives $m\dot{\mathbf{v}} = 0$, and hence $\mathbf{v}$ is a constant. The solution thus corresponds to uniform motion, and there is no loss of energy by radiation.

(ii) *Uniform electric field*

For an electron in a uniform electric field $\mathbf{E}_0$, the equation of motion (9) gives

$$m\dot{\mathbf{v}} = \left(1+\frac{D}{a}\right)e\mathbf{E}_0 = e\mathbf{E}_0.$$

The solution corresponds to the usual motion in a parabola; and according to (15) the rate of loss of energy is $\mathscr{R} = \frac{2}{3}e^4\mathbf{E}_0^2/m^2$.

(iii) *Scattering of light*

Consider the rectilinear motion of an electron in the field of an incident beam of light of frequency ν whose electric field components are

$$E_x = \mathscr{E}_0 \cos \nu t, \quad E_y = E_z = 0. \tag{16}$$

We suppose that $\mathscr{E}_0$ is small and that the frequency ν is not larger than a. Then the electron's velocity and higher derivatives are small enough for products of two or more of them to be negligible, and the nth derivative of $\mathbf{E}$ tends to zero as n tends to infinity. The equations of motion (9) are then applicable. We obtain

$$m\dot{v} = \left(1+\frac{D}{a}\right)e\mathscr{E}_0 \cos \nu t = e\mathscr{E}_0\left(\cos \nu t - \frac{\nu}{a}\sin \nu t\right); \tag{17}$$

that is,

$$\ddot{x} = A \cos(\nu t + \alpha), \tag{18}$$

where

$$A = \frac{e\mathscr{E}_0}{m}\left(1+\frac{\nu^2}{a^2}\right)^{\frac{1}{2}}, \quad \tan\alpha = \frac{\nu}{a} \quad (0<\alpha<\tfrac{1}{2}\pi). \tag{19}$$

Hence

$$x = -\frac{A}{\nu^2}\cos(\nu t+\alpha). \tag{20}$$

The electron performs a vibration with the same frequency as the incident wave and with a phase difference $\pi+\alpha$:

$$v^{(2n)} = \frac{A}{\nu}(-\nu^2)^n \sin(\nu t+\alpha), \quad v^{(2n+1)} = \frac{A}{\nu}(-\nu^2)^n \nu \cos(\nu t+\alpha).$$

The rate of loss of energy according to (15) is

$$\mathscr{R} = \frac{mA^2}{a}\sum_{n=1}^{\infty}\left\{\frac{\nu^{4n-4}}{a^{4n-4}}\cos^2(\nu t+\alpha) - \frac{\nu^{4n-2}}{a^{4n-2}}\sin^2(\nu t+\alpha)\right\}. \tag{21}$$

Hence the average rate of loss of energy is

$$\begin{aligned}\overline{\mathscr{R}} &= \frac{1}{2}\frac{mA^2}{a}\sum_{n=1}^{\infty}\left\{\frac{\nu^{4n-4}}{a^{4n-4}} - \frac{\nu^{4n-2}}{a^{4n-2}}\right\} = \frac{1}{2}\frac{mA^2}{a}\sum_{n=1}^{\infty}(-1)^{n-1}\left(\frac{\nu^2}{a^2}\right)^{n-1} \\ &= \frac{1}{2}\frac{mA^2}{a}\frac{1}{1+\nu^2/a^2} = \frac{1}{3}\frac{e^4\mathscr{E}_0^2}{m^2}.\end{aligned} \tag{22}$$

This expression for $\overline{\mathscr{R}}$ is the same as the classical Thomson formula for the average rate of energy loss. It differs from the corresponding expression obtained when one uses the Lorentz-Dirac equations, according to which the rate of energy loss is found to be

$$\overline{\mathscr{R}} = \frac{1}{3}\frac{e^4\mathscr{E}_0^2}{m^2(1+\nu^2/a^2)}. \tag{23}$$

We have here a point of difference between the theory we are using in this paper and the Dirac classical theory of a point electron. The difference, however, is at most by a factor $\frac{1}{2}$ in the range of frequencies we are considering here, namely, $\nu \leqslant a$.

(iv) *Harmonic oscillator*

Suppose the binding force of the oscillator is $m\omega^2 x$. The equation of motion (9) gives

$$m\ddot{x} = -\left(1+\frac{D}{a}\right)m\omega^2 x,$$

that is,

$$\ddot{x}+\frac{\omega^2}{a}\dot{x}+\omega^2 x = 0. \tag{24}$$

These equations are valid only when ω is small, and then the solution is of the form

$$x = e^{-\omega^2 t/2a}\left[C\cos\left\{\omega\left(1-\frac{\omega^2}{4a^2}\right)^{\frac{1}{2}}t\right\}+D\sin\left\{\omega\left(1-\frac{\omega^2}{4a^2}\right)^{\frac{1}{2}}t\right\}\right], \tag{25}$$

where C and D are arbitrary constants, showing that the motion is approximately periodic with the oscillations gradually dying out. The lifetime of the oscillator and the line breadth may be calculated in the usual way.

(v) *Scattering of light by an oscillator*

Consider the rectilinear motion of an oscillator with binding force $m\omega^2 x$, placed in the field of an incident beam of light with electric field components given by (16). The equation of motion is

$$m\ddot{x} = \left(1+\frac{D}{a}\right)(-m\omega^2 x + e\mathscr{E}_0\cos\nu t),$$

that is,

$$\ddot{x}+\frac{\omega^2}{a}\dot{x}+\omega^2 x = A\cos(\nu t+\alpha), \tag{26}$$

where A and α are as in (19). Ignoring the complementary function which tends to zero exponentially with time, we take as the solution

$$x = B\cos(\nu t+\alpha+\beta), \tag{27}$$

where

$$B = A\bigg/\left\{(\nu^2-\omega^2)^2+\frac{\omega^4\nu^2}{a^2}\right\}^{\frac{1}{2}}, \quad \tan\beta = \frac{\omega^2\nu}{a(\omega^2-\nu^2)}. \tag{28}$$

The average rate of energy loss is then

$$\overline{\mathscr{R}} = \frac{1}{2}\frac{mB^2\nu^4}{a(1+\nu^2/a^2)},$$

as in (22). Substituting for B we obtain

$$\overline{\mathscr{R}} = \frac{e^4\mathscr{E}_0^2\nu^4}{3m^2[(\nu^2-\omega^2)^2+\omega^4\nu^2/a^2]}. \tag{29}$$

From (29) we easily derive the dispersion formula.

6. HIGHER DERIVATIVES IN DIRAC'S THEORY

The higher derivatives of velocity may also be introduced into the equations of motion in Dirac's theory. I have considered in a previous paper equations of motion containing the second-order derivatives of velocity (Eliezer 1946). We shall now proceed to investigate the form of the equations of motion containing other higher derivatives. It appears that only the even higher derivatives of velocity occur in the equations of motion. This is because Dirac's theory uses a point model of the electron, and therefore the rate of energy loss is given by the usual expression for the rate of energy loss of a point charge, which in the non-relativistic limit is $\frac{2}{3}e^2\dot{\mathbf{v}}^2$. If odd derivatives of velocity appear in the equations of motion they will give rise to irreversible loss of energy of the electron in a way not consistent with conservation requirements for a point electron.

We shall first consider the relativistic equations of motion. Let z_μ $(\mu = 0, 1, 2, 3)$ denote the space-time co-ordinates of the electron; dots now denote differentiation with respect to the proper time s, and $v_\mu = \dot{z}_\mu$. Then v_μ satisfies the condition

$$v^\mu v_\mu \equiv v^2 = 1. \tag{30}$$

The equations of motion of an electron in an electromagnetic field described by the field quantities $f_{\mu\nu}$ is

$$\tfrac{2}{3}e^2(\ddot{v}_\mu + \dot{v}^2 v_\mu) - ev^\sigma f_{\mu\sigma} = \dot{B}_\mu, \tag{31}$$

where B_μ is an arbitrary function depending on the particle's velocity and higher derivatives, and satisfying the requirements

$$(v, \dot{B}) = 0, \tag{32}$$

and

$$v_\lambda B_\mu - v_\mu B_\lambda \tag{33}$$

must be a perfect differential with respect to the proper time s (Bhabha 1939). The scalar product notation

$$(a, b) \equiv a^\mu b_\mu = a_0 b_0 - a_1 b_1 - a_2 b_2 - a_3 b_3$$

is used. Let $v_\mu^{(n)}$ denote the nth derivative of v_μ with respect to the proper time. We solve for B_μ by assuming it to be expressible as a function of the form

$$B_\mu = B_0 v_\mu + B_1 v_\mu^{(1)} + \ldots + B_n v_\mu^{(n)} + \ldots, \tag{34}$$

where $B_0, B_1, \ldots$ are scalar quantities which are invariant functions of the velocity components and all their derivatives. The condition (32) gives

$$\dot{B}_0 + (B_1 + \dot{B}_2)(v, v^{(2)}) + \ldots + (B_{n-1} + \dot{B}_n)(v, v^{(n)}) + \ldots = 0. \tag{35}$$

We note that

$$(v, v^{(2n)}) = \frac{d}{ds}[(v, v^{(2n-1)}) - (v^{(1)} v^{(2n-2)}) + \ldots + (-1)^{n-1}(v^{(n-1)}, v^{(n)})] + \tfrac{1}{2}(-1)^n v^{(n)2}, \tag{36}$$

$$(v, v^{(2n+1)}) = \frac{d}{ds}[(v, v^{(2n)}) - (v^{(1)}, v^{(2n-1)}) + \ldots + (-1)^n (v^{(n-1)}, v^{(n+1)}) + \tfrac{1}{2}(-1)^n v^{(n)2}]. \tag{37}$$

The condition (33) requires that

$$B_1(v_\lambda v_\mu^{(1)} - v_\mu v_\lambda^{(1)}) + \ldots + B_n(v_\lambda v_\mu^{(n)} - v_\mu v_\lambda^{(n)}) + \ldots \tag{38}$$

must be a perfect differential. But

$$v_\lambda v_\mu^{(2n)} - v_\mu v_\lambda^{(2n)} = \frac{d}{ds}[\{v_\lambda v_\mu^{(2n-1)} - v_\mu v_\lambda^{(2n-1)}\} - \ldots + (-1)^n \{v_\lambda^{(n-1)} v_\mu^{(n)} - v_\mu^{(n-1)} v_\lambda^{(n)}], \tag{39}$$

$$\begin{aligned} v_\lambda v_\mu^{(2n+1)} - v_\mu v_\lambda^{(2n+1)} = \frac{d}{ds}[\{v_\lambda v_\mu^{(2n)} - v_\mu v_\lambda^{(2n)}\} - \ldots \\ + (-1)^{n-1}\{v_\lambda^{(n-1)} v_\mu^{(n+1)} - v_\mu^{(n)} v_\lambda^{(n+1)}\}] + (-1)^n \{v_\lambda^{(n)} v_\mu^{(n+1)} - v_\mu^{(n)} v_\lambda^{(n+1)}\}. \end{aligned} \tag{40}$$

On examining the conditions (35) and (38) we see that if we choose the B's so that

$$B_1 = B_3 = \ldots = B_{2n+1} = \ldots = 0, \quad B_2, B_4, \ldots, B_{2n}, \ldots \text{ are constants}, \tag{41}$$

then (38) is automatically satisfied, and (35) gives

$$\dot{B}_0 + B_2(v, v^{(3)}) + \ldots + B_{2n}(v, v^{(2n+1)}) + \ldots = 0. \tag{42}$$

Integrating, we obtain

$$B_0 = m - \sum_{n=1}^{\infty} B_{2n}[(v, v^{(2n)}) - (v^{(1)}, v^{(2n-1)}) + \ldots + (-1)^n (v^{(n-1)}, v^{(n+1)}) + \tfrac{1}{2}(-1)^n v^{(n)^2}], \tag{43}$$

where m is an arbitrary constant of integration. Therefore we take

$$B_\mu = mv_\mu + \sum_{n=1}^{\infty} B_{2n}[v_\mu^{(2n)} - \{(v, v^{(2n)}) - \ldots + (-1)^n (v^{(n-1)}, v^{(n+1)}) + \tfrac{1}{2}(-1)^n v^{(n)^2}\} v_\mu]. \tag{44}$$

This choice of B_μ leads to the equations of motion

$$\begin{aligned} m\dot{v}_\mu - \tfrac{2}{3}e^2(\ddot{v}_\mu + \dot{v}^2 v_\mu) + \sum_{n=1}^{\infty} B_{2n}[v_\mu^{(2n+1)} - (v, v^{(2n+1)})\, v_\mu - \{(v, v^{(2n)}) - \ldots \\ + (-1)^n (v^{(n-1)}, v^{(n+1)}) + \tfrac{1}{2}(-1)^n v^{(n)^2}\} \dot{v}_\mu] = ev^\sigma f_{\mu\sigma}, \end{aligned} \tag{45}$$

where the constants $B_2, B_4, \ldots$ are arbitrary. The constant m which appears here as an arbitrary constant may be identified with the mass of the electron.

Suppose we now consider a motion in which the electron's velocity and all its derivatives are small, so that products of two or more of them may be neglected. If $\mathbf{v}$ now denotes $d\mathbf{z}/dt$, the equations of motion of an electron in an electric field $\mathbf{E}$ become

$$m\frac{d\mathbf{v}}{dt} - \tfrac{2}{3}e^2 \frac{d^2\mathbf{v}}{dt^2} + B_2 \frac{d^3\mathbf{v}}{dt^3} + \ldots + B_{2n} \frac{d^{2n+1}\mathbf{v}}{dt^{2n+1}} + \ldots = e\mathbf{E}, \tag{46}$$

that is,

$$\left[1 - \frac{D}{a} + a_2 D^2 + a_4 D^4 + \ldots + a_{2n} D^{2n} + \ldots\right] m \frac{d\mathbf{v}}{dt} = e\mathbf{E}, \tag{47}$$

where $D \equiv d/dt$, $a = 3m/2e^2$ and $a_{2n} = B_{2n}/m$.

The series

$$1 - \frac{x}{a} + a_2 x^2 + a_4 x^4 + \ldots + a_{2n} x^{2n} + \ldots \tag{48}$$

has an inverse for sufficiently small values of x. Suppose the inverse is

$$1 + \frac{x}{a} + c_2 x^2 + c_3 x^3 + \ldots + c_r x^r + \ldots, \tag{49}$$

where the constants $c_2, c_3, \dots$ are determined in terms of $a_2, a_4, \dots$, that is, in terms of the arbitrary constants $B_2, B_4, \dots$. Therefore, analogous to (9) we would now have

$$m\frac{d\mathbf{v}}{dt} = \left(1+\frac{D}{a}+c_2D^2+\dots+c_rD^r+\dots\right)e\mathbf{E}. \tag{50}$$

In the particular case of the Lorentz-Dirac equations of motion, the constants $B_2, B_4, \dots$ are all zero, and the equations corresponding to (50) are then

$$m\frac{d\mathbf{v}}{dt} = \left(1+\frac{D}{a}+\frac{D^2}{a^2}+\dots+\frac{D^r}{a^r}+\dots\right)e\mathbf{E}. \tag{51}$$

If a finite model of the electron is used, and the same procedure as in the above be followed, with allowance made for the variation of the external field within the interior of the electron, then analogous to (9) the equations describing the motion of the centre of inertia of the electron may be expected to have the same form as (50). The first two terms on the right-hand side, namely $e\mathbf{E}$ and $a^{-1}e\dot{\mathbf{E}}$, are independent of the structure of the electron, while the remaining terms of the series all depend on the charge distribution. It seems therefore plausible to take the equations

$$m\frac{d\mathbf{v}}{dt} = \left(1+\frac{D}{a}\right)e\mathbf{E} \tag{9}$$

as of special physical significance.

An interesting point to note is that the familiar procedure of solving the Lorentz equations (1) approximately by first neglecting the term $\frac{2}{3}e^2\ddot{\mathbf{v}}$ altogether, and then carrying the solution to a higher approximation by replacing $\frac{2}{3}e^2\ddot{\mathbf{v}}$ by its value according to the first approximate solution leads precisely to the equations (9). For if in the first approximation $m\dot{\mathbf{v}} = e\mathbf{E}$, then in the second approximation

$$m\dot{\mathbf{v}} = e\mathbf{E}+\tfrac{2}{3}e^2\ddot{\mathbf{v}} = \left(1+\frac{D}{a}\right)e\mathbf{E}.$$

In our theory, however, we shall be taking the equations (9) as exact equations within certain limits.

7. RELATIVISTIC EQUATIONS

The equations of motion (9) have been derived from the Maxwell-Lorentz theory. According to this derivation these equations are valid only when the velocity and all its derivatives are small so that products of two or more of them may be neglected, and the external field does not vary too rapidly. Now we shall seek a set of simple relativistic equations which under the above conditions would reduce to the equations (9), and which would be free from self-accelerating motions. Using the notation of § 6, we see that the equations

$$m\dot{v}_\mu = ev^\sigma f_{\mu\sigma}+\frac{e}{a}\left\{\frac{d}{ds}(v^\sigma f_{\mu\sigma})-\dot{v}^\sigma v^\rho f_{\rho\sigma}v_\mu\right\} \tag{52}$$

satisfy this requirement. On taking the scalar product of the equation (52) with v^μ we see that both sides give zero, owing to the condition $v^2=1$ and the antisymmetry of $f_{\mu\sigma}$. The last term in curly brackets in (52) gives the contribution from the radiation reaction.

We shall now postulate a classical theory in which the motion of an electron in a given electromagnetic field is described exactly by the equations (52), these equations being now looked upon as exact equations of motion applicable under all conditions within the limits of the classical theory. In taking (52) as the exact equations of motion, one should expect that for high velocities and rapidly varying fields there would be departures from the Maxwell-Lorentz theory. Maxwell's equations for empty space will be assumed to hold, that is,

$$\frac{\partial f_{\mu\sigma}}{\partial x^\rho}+\frac{\partial f_{\sigma\rho}}{\partial x^\mu}+\frac{\partial f_{\rho\mu}}{\partial x^\sigma}=0. \tag{53}$$

The energy equation, obtained from (52) with $\mu = 0$, is

$$m\dot{v}_0-ev^\sigma f_{0\sigma}=\frac{e}{a}\left\{\frac{d}{ds}(v^\sigma f_{0\sigma})-\dot{v}^\sigma v^\rho f_{\rho\sigma}v_0\right\}. \tag{54}$$

The first term on the left-hand side gives the rate of increase of kinetic energy and the second term the rate of increase of potential energy. The right-hand side gives the rate of change of the total energy.

We shall consider some applications of the equations (52). The solution of the equations of motion contains the same number of arbitrary constants of integration as in the elementary theory which ignores radiation damping. The equations (52) do not have solutions corresponding to self-accelerating motions.

Let us investigate the cross-section for the scattering of light by an electron on the basis of the equations (52). Let us suppose that the amplitude $\mathscr{E}_0$ of the incident wave of light is so small that the velocity acquired by the electron is small enough for the non-relativistic equations to be valid. The motion is given by

$$m\dot{v}=e\mathscr{E}_0\left(\cos\nu t-\frac{\nu}{a}\sin\nu t\right),$$

which is the same as (17) except that this equation is now supposed to be valid for all values of the frequency ν, whereas previously there was the restriction $\nu\leqslant a$. The rate of energy loss is found to be

$$\begin{aligned}\mathscr{R}&=-\frac{1}{a}\dot{E}v=-\frac{e}{a}(-\mathscr{E}_0\nu\sin\nu t)\frac{e\mathscr{E}_0}{m}\left(\frac{\sin\nu t}{\nu}+\frac{\cos\nu t}{a}\right)\\&=\frac{2}{3}\frac{e^4\mathscr{E}_0^2}{m^2}\left(\sin^2\nu t+\frac{\nu}{a}\sin\nu t\,\cos\nu t\right).\end{aligned} \tag{55}$$

Hence the average value of $\mathscr{R}$ over a period is

$$\overline{\mathscr{R}}=\tfrac{1}{3}e^4\mathscr{E}_0^2/m^2,$$

which agrees with the previous result. This shows that on this theory the total cross-section for the scattering of light is independent of the frequency and is a universal constant. This shows a difference from the Dirac theory, where the rate of energy loss is given by (23) and according to which the scattering cross-section varies as ν^{-2} for large frequencies.

Similar calculations may be made for the harmonic oscillator. The equations of motion are the same as (24), which for $\omega < 2a$ has solution of the form (25). If $\omega \geqslant 2a$ the auxiliary roots are real and negative, and the corresponding solution is such that the electron rapidly comes to rest without oscillating. For the scattering of light by the oscillator the average rate of energy loss $\overline{\mathscr{R}}$ is found to be given by the formula (29), now supposed to be valid for all values of ω and ν. When $\omega \geqslant a\sqrt{2}$ the expression $\nu^4 \Big/ \left\{(\nu^2-\omega^2)^2+\frac{\omega^4\nu^2}{a^2}\right\}$ lies between 0 and 1 for all values of ν, while if $\omega < a\sqrt{2}$ the expression has a maximum value $4a^4/\omega^2(4a^2-\omega^2)$ when $\nu = a\omega \Big/ \left(a^2-\frac{\omega^2}{2}\right)^{\frac{1}{2}}$. This shows that when ω is small $\overline{\mathscr{R}}$ has a large value for frequencies near ω, giving an example of resonance fluorescence.

The hydrogen atom. As a further application of the relativistic equations of motion, we shall consider the rectilinear motion of an electron towards a proton which we shall take to be fixed by regarding its mass as infinitely large. If t, x denote the time and space co-ordinates of the electron the equations of motion give

$$m\ddot{x} = -e^2\left(\frac{1}{x^2}-\frac{2\dot{x}}{ax^3}\right)\dot{t}, \quad \dot{t}^2-\dot{x}^2 = 1, \tag{56}$$

where dots denote the differentiation with respect to the proper time s. Writing $V = -\dot{x}$, substituting for $\dot{t}$ and transforming the independent variable to x, we obtain

$$V\frac{dV}{dx} = -\frac{3}{2a}\left(\frac{1}{x^2}+\frac{2V}{ax^3}\right)(1+V^2)^{\frac{1}{2}}. \tag{57}$$

If initially the electron is projected from a point $x = x_0 > 0$ towards the proton, then V is positive and hence dV/dx is negative. Therefore V increases as x decreases. That is, the electron approaches the proton with increasing speed and eventually collides with it (unlike the motion as given by the Lorentz-Dirac equations, according to which the electron is brought to rest before it could reach the proton). An approximate solution for small values of x is seen to be

$$V \sim C\exp\left(\frac{3}{2a^2x^2}\right), \tag{58}$$

where C is a positive constant. The solution seems to be in agreement with the usual results of mechanics.

There is another equation of motion which one may take as the starting point of a classical theory, and which is somewhat simpler than the equation (52), namely,

$$m\dot{v}_\mu = ev^\sigma\left(f_{\mu\sigma}+\frac{1}{a}\dot{f}_{\mu\sigma}\right). \tag{59}$$

For rectilinear motion in an electric field (52) and (59) are equivalent. But in other cases the equations (59) leads to some fundamental differences from the Maxwell-Lorentz theory—for example, an electron moving in a magnetic field does not lose energy by radiation, according to (59).

8. DISCUSSION

The equations of motion (52) are seen to have many satisfactory features within the limits of the classical theory. They are free from self-accelerating motions. For small velocities and for fields which do not vary rapidly, these equations give the same results as the Maxwell-Lorentz theory; and in other cases they appear to give results in harmony with the usual notions of mechanics. A number of interesting points require further investigation. The extent to which Maxwell's equations need modification to be applicable when charged matter is present, the interaction of two moving electrons, the two-dimensional motion of the electron in the hydrogen atom, the passage to the quantum theory—all these questions require much detailed work, and it is hoped to deal with these in a later paper.

I record with gratitude my thanks to Professor P. A. M. Dirac for helpful discussions, to Professor R. E. Peierls for suggestions on the writing of this paper, and to the Royal Commissioners for the Exhibition of 1851 for the award of a senior studentship.

REFERENCES

Bhabha, H. J. 1939 *Proc. Indian Acad. Sci.* A, **10**, 324.
Dirac, P. A. M. 1938 *Proc. Roy. Soc.* A, **167**, 148.
Dirac, P. A. M. 1946 *Commun. Dublin Inst.* A, **3**, 1.
Eliezer, C. J. 1943 *Proc. Camb. Phil. Soc.* **39**, 173.
Eliezer, C. J. 1945 *Bull. Calcutta Math. Soc.* **37**, 125.
Eliezer, C. J. 1946 *Proc. Camb. Phil. Soc.* **42**, 40.
Eliezer, C. J. 1947 *Rev. Mod. Phys.* **19**, 147.
Eliezer, C. J. & Mailvaganam, A. W. 1945 *Proc. Camb. Phil. Soc.* **41**, 184.
Lorentz, H. A. 1909 *The theory of electrons.* Leipzig: Teubner.
Page, L. 1918 *Phys. Rev.* **9**, 376.
Wheeler, J. A. & Feynman, R. P. 1945 *Rev. Mod. Phys.* **17**, 157.

Adjustments in bacterial reaction systems

I. The reducing power of *Bact. lactis aerogenes* under various conditions

By P. R. Lewis and C. N. Hinshelwood, F.R.S.

Physical Chemistry Laboratory, University of Oxford

(*Received* 31 *March* 1948)

ABSTRACT

The reducing power of cells of *Bact. lactis aerogenes* has been measured by dye reduction tests with resting and with growing cultures in media containing as nitrogen sources ammonium salts or nitrate, and under aerobic and anaerobic conditions.

Changes in growth rate have been correlated with changes in the reducing power accompanying transitions from one medium to the other or from anaerobic to aerobic conditions and vice versa.

Two types of adjustment play a part in the observed phenomena: (*a*) rapid changes in concentrations of active intermediates, (*b*) slow adaptive modifications of enzyme systems in response to needs of the cell.

The former are exemplified *inter alia* by the rapid fall in reducing power and growth rate when an anaerobic nitrate culture is aerated. The latter include the development of extra reducing power when the cell has to use nitrate as a source of nitrogen and of still more when the nitrate has to be reduced in presence of oxygen.

Adjustments in bacterial reaction systems

II. Adaptive mechanisms

By C. N. Hinshelwood, F.R.S. and P. R. Lewis

Physical Chemistry Laboratory, University of Oxford

(*Received* 31 *March* 1948)

ABSTRACT

General propositions about the way in which enzyme systems can adapt themselves to respond to new needs of bacterial cells are considered.

In certain conditions the establishment of an optimum growth rate is automatic. In the simplest treatment the 'enzyme' material is regarded as autosynthetic, but it is shown that similar equations may hold where the synthesis of one enzyme is linked with the working of another enzyme rather than with its own.

(The full texts of these two papers are published in *Proc. Roy. Soc.* Series B, No. 880 (Vol. **135**, pp. 301–322)

INDEX TO VOLUME 194 (A)

Adjustments in bacterial reaction systems. I. Reducing power of *Bact. lactis aerogenes* under various conditions. II. Adaptive mechanisms. (Abstracts) Lewis & Hinshelwood), 556.

Allan, H. R. & Wilkinson, C. A. The transmutation of magnesium, aluminium and silicon by deuterons, 131.

Andrew, E. R. The intermediate state of superconductors. II. The resistance of cylindrical superconductors in transverse magnetic fields, 80.

Andrew, E. R. The intermediate state of superconductors. III. Theory of behaviour of superconducting cylinders in transverse magnetic fields, 98.

Ashauer, Sonja. A generalization of the method of separating longitudinal and transverse waves in electrodynamics, 206.

Aspects of a theory of unimolecular reaction rates (Slater), 112.

Atmospheric tides in the ionosphere (Martyn). III. Lunar tidal variations at Canberra, 429. IV. Studies of the solar tide, and the location of the regions producing the diurnal magnetic variations, 445.

Batchelor, G. K. & Townsend, A. A. Decay of turbulence in the final period, 527.

Bhatia, A. B. & Krishnan, Sir K. S. 'Diffuse scattering' of the Fermi electrons in monovalent metals in relation to their electrical resistivities, 185.

Binnie, A. M. & Hookings, G. A. Laboratory experiments on whirlpools, 398.

Blackman, M., Egerton, Sir Alfred & Truter, E. V. Heat transfer by radiation to surfaces at low temperatures, 147.

Bloom, H., Harrap, B. S. & Heymann, E. Determination of the viscosity of molten salts by a capillary method. The system cadmium chloride-cadmium bromide, 237.

Bullard, E. C. & Cooper, R. I. B. The determination of the masses necessary to produce a given gravitational field, 332.

Calculation of surface tension from measurements of pendant drops (Fordham), 1.

Carrington, W. E. & Gayler, Marie L. V. The use of flat-ended projectiles for determining dynamic yield stress. III. Changes in microstructure caused by deformation under impact at high-striking velocities, 323.

Chalkin, F. C. Intensity measurements in the very soft X-ray region, 42.

Classical theory of particles (Eliezer), 543.

Comparative X-ray study of foetal and adult sheep haemoglobins (Kendrew & Perutz), 375.

Constitution of the aluminium-rich aluminium-cobalt-iron alloys, with reference to the role of transitional elements in alloy formation (Raynor & Waldron), 362.

Cooper, R. I. B. *See* Bullard & Cooper.

Davies, J. T. & Rideal, E. K. The kinetic salt effect in monolayer reactions, 417.

Decay of turbulence in the final period (Batchelor & Townsend), 527.

Désirant, M. & Shoenberg, D. The intermediate state of superconductors. I. Magnetization of superconducting cylinders in transverse magnetic fields, 63.

Determination of the masses necessary to produce a given gravitational field (Bullard & Cooper), 332.

Determination of the viscosity of molten salts by a capillary method. The system cadmium chloride-cadmium bromide (Bloom, Harrap & Heymann), 237.

Detonation of solid explosives: the equilibrium conditions in the detonation wave-front and the adiabatic expansion of the products of detonation (Jones & Miller), 480.

'Diffuse scattering' of the Fermi electrons in monovalent metals in relation to their electrical resistivities (Bhatia & Krishnan), 185.

Effect of compressibility on laminar boundary layers and their separation (Howarth), 16.

Egerton, Sir Alfred. *See* Blackman, Egerton & Truter.

Electroviscosity. I. Flow of liquids between surfaces in close proximity, 259. II. Experimental demonstration of the electroviscous effect (Elton), 275.

Eliezer, C. J. On the classical theory of particles, 543.

Elton, G. A. H. Electroviscosity. I. The flow of liquids between surfaces in close proximity, 259. II. Experimental demonstration of the electroviscous effect, 275.

Essen, L. & Gordon-Smith, A. C. The velocity of propagation of electromagnetic waves derived from the resonant frequencies of a cylindrical cavity resonator, 348.

Fordham, S. On the calculation of surface tension from measurements of pendant drops, 1.

Gaydon, A. G. & Wolfhard, H. G. Spectroscopic studies of low-pressure flames; temperature measurements in acetylene flames, 169.

Gayler, Marie L. V. *See* Carrington & Gayler.

General kinetic theory of liquids. V. Liquid He II (Green), 244.

Generalization of the method of separating longitudinal and transverse waves in electrodynamics (Ashauer), 206.

Gordon-Smith, A. C. *See* Essen & Gordon-Smith.

Green, H. S. A general kinetic theory of liquids. V. Liquid He II, 244.

Harrap, B. S. *See* Bloom, Harrap & Heymann.

Heat transfer by radiation to surfaces at low temperatures (Blackman, Egerton & Truter), 147.

Heymann, E. *See* Bloom, Harrap & Heymann.

Hinshelwood, C. N. *See* Lewis & Hinshelwood (Abstracts).

Hookings, G. A. *See* Binnie & Hookings.

Howarth, L. Concerning the effect of compressibility on laminar boundary layers and their separation, 16.

Intensity measurements in the very soft X-ray region (Chalklin), 42.

Intermediate state of superconductors. I. Magnetization of superconducting cylinders in transverse magnetic fields (Désirant & Shoenberg), 63. II. Resistance of cylindrical superconductors in transverse magnetic fields (Andrew), 80. III. Theory of behaviour of superconducting cylinders in transverse magnetic fields (Andrew), 98.

Jones, H. & Miller, A. R. The detonation of solid explosives: the equilibrium conditions in the detonation wave-front and the adiabatic expansion of the products of detonation, 480.

Kendrew, J. C. & Perutz, M. F. A comparative X-ray study of foetal and adult sheep haemoglobins, 375.

Kinetic salt effect in monolayer reactions (Davies & Rideal), 417.

Krishnan, Sir K. S. *See* Bhatia & Krishnan.

Laboratory experiments on whirlpools (Binnie & Hookings), 398.

Laminar boundary-layer equations of bodies of revolution. Motion of a sphere (Meksyn), 218.

Lewis, P. R. & Hinshelwood, C. N. Adjustments in bacterial reaction systems. I. The reducing power of *Bact. lactis aerogenes* under various conditions. II. Adaptive mechanisms (Abstracts), 556.

Martyn, D. F. Atmospheric tides in the ionosphere. III. Lunar tidal variations at Canberra, 429. IV. Studies of the solar tide, and the location of the regions producing the diurnal magnetic variations, 445.

Meksyn, D. The laminar boundary-layer equations of bodies of revolution. Motion of a sphere, 218.

Miller, A. R. *See* Jones & Miller.

Perutz, M. F. *See* Kendrew & Perutz.

Ramsey, W. H. On the scattering of protons by protons, 228.
Raynor, G. V. & Waldron, M. B. The constitution of the aluminium-rich aluminium-cobalt-iron alloys, with reference to the role of transitional elements in alloy formation, 362.
Rideal, E. K. *See* Davies & Rideal.

Scattering of protons by protons (Ramsey), 228.
Shoenberg, D. *See* Désirant & Shoenberg.
Silver, R. S. Temperature and pressure phenomena in the flow of saturated liquids, 464.
Slater, N. B. Aspects of a theory of unimolecular reaction rates, 112.
Spectroscopic studies of low-pressure flames; temperature measurements in acetylene flames (Gaydon & Wolfhard), 169.
Study of expected loss rates in the counting of particles from pulsed sources (Westcott), 508.

Taylor, Sir Geoffrey. The use of flat-ended projectiles for determining dynamic yield stress. I. Theoretical considerations, 289.
Temperature and pressure phenomena in the flow of saturated liquids (Silver), 464.
Townsend, A. A. *See* Batchelor & Townsend.
Transmutation of magnesium, aluminium and silicon by deuterons (Allan & Wilkinson), 131.
Truter, E. V. *See* Blackman, Egerton & Truter.

Use of flat-ended projectiles for determining dynamic yield stress. I. Theoretical considerations (Taylor), 289. II. Tests on various metallic materials (Whiffin), 300. III. Changes in microstructure caused by deformation under impact at high-striking velocities (Carrington & Gayler), 323.

Velocity of propagation of electromagnetic waves derived from the resonant frequencies of a cylindrical cavity resonator (Essen & Gordon-Smith), 348.

Waldron, M. B. *See* Raynor & Waldron.
Westcott, C. H. A study of expected loss rates in the counting of particles from pulsed sources, 508.
Whiffin, A. C. The use of flat-ended projectiles for determining dynamic yield stress. II. Tests on various metallic materials, 300.
Wilkinson, Christiane A. *See* Allan & Wilkinson.
Wolfhard, H. G. *See* Gaydon & Wolfhard.

END OF THE ONE HUNDRED AND NINETY-FOURTH VOLUME (SERIES A)

Indian Agricultural Research Institute Library, New Delhi

1. Books are issued for 14 days, beyond it for next 7 days a fine of 15 paise & after words 25 paise per day shall be charged.
2. Books may be renewed at the discretion of the head Library Services.
3. Dog-earing of pages of a book, marking of writing there in with ink or pencil, tearing or taking out its pages or otherwise damaging it will constitute an injury to a book.
4. Unless a borrower points out the injury at the time of borrowing the book, he shall be required to replace the book or pay its price, if detected at the time of return.

HELP US TO KEEP THIS BOOK FRESH AND CLEAN

I.A.R.I. 75

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY, NEW DELHI.

Date of Issue	Date of Issue	Date of Issue
I\1	15\2\90	

GIPNLK—H-40 I.A.R.I.—29-4-55—15,000